Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3902~3910
e-ISSN: 2548-6861

Sentiment Analysis of Skincare Product Reviews: A Comparison of Naïve Bayes and Support Vector Machine Algorithms

Refida Septiana Putri [1]*, Reykha Putri Randika [2]*, Febi Dwi Sasmita[3]*, Pungkas Subarkah[4]*
* Informatika, Universitas Amikom Purwokerto
refidaseptianaputri0@gmail.com [1], reykhaputri690@gmail.com [2], sasmitafebidwi@gmail.com [3], subarkah@amikompurwokerto.ac.id [4]

Article Info

Article history:

Received 2026-06-05
Revised 2026-07-02
Accepted 2026-08-12

Keyword:

Sentiment Analysis,
Multinomial Naïve Bayes,
Support Vector Machine,
Skincare Reviews,
Class Imbalance.

ABSTRACT

The rapid growth of the skincare industry has generated a massive volume of consumer reviews on e-commerce platforms, making manual sentiment analysis increasingly impractical. This study compares the performance of Multinomial Naïve Bayes and Support Vector Machine (SVM) for sentiment classification of skincare product reviews using the Sephora Products and Skincare Reviews dataset from Kaggle, consisting of 602,130 reviews. Unlike previous studies that often employ different datasets and experimental settings, this research evaluates both algorithms using a uniform pipeline, including automatic sentiment labeling based on rating values, text preprocessing, Bag of Words feature representation, and identical evaluation procedures. Model performance was assessed using Area Under Curve (AUC), Accuracy, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC). The results indicate that positive sentiment dominates the dataset (82.37%), resulting in a highly imbalanced class distribution. Multinomial Naïve Bayes achieved better performance than SVM on most evaluation metrics, with an AUC of 0.784, F1-Score of 0.764, Precision of 0.749, and MCC of 0.153, whereas SVM obtained an AUC of 0.497, F1-Score of 0.743, Precision of 0.695, and MCC of 0.000. The near-zero MCC and low AUC of SVM suggest that the model struggled to distinguish minority classes under extreme class imbalance despite achieving high accuracy. These findings highlight the importance of employing multiple evaluation metrics beyond accuracy when assessing classification performance on imbalanced datasets. Furthermore, the results indicate that Multinomial Naïve Bayes provided better performance than SVM under the dataset characteristics and experimental configuration used in this study.

I. PENDAHULUAN

Industri perawatan kulit (*skincare*) saat ini tengah mengalami pertumbuhan yang sangat pesat, sejalan dengan meningkatnya kesadaran masyarakat akan pentingnya kesehatan kulit serta perkembangan platform digital yang mendukung transaksi produk[1][2]. Platform *e-commerce* menyediakan berbagai variasi produk *skincare* seperti pembersih wajah (*cleansers*), toner, pelembab (*moisturizers*), serum, hingga tabir surya (*sunscreens*), yang disertai dengan akumulasi ulasan konsumen dalam jumlah yang sangat besar [1] . Ulasan tersebut menjadi sumber informasi yang penting bagi calon konsumen karena memuat pengalaman langsung pengguna dengan produk tersebut[3]. Namun, volume ulasan yang terus meningkat secara berkelanjutan membuat proses analisis ulasan secara manual menjadi tidak efisien [4].

Untuk mengatasi tantangan ini, analisis sentimen dapat diterapkan sebagai pendekatan otomatis untuk memproses dan mengklasifikasikan volume ulasan yang besar secara efektif. Analisis sentimen adalah cabang dari *text mining* yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam data teks ke dalam kategori tertentu seperti positif, negatif, atau netral [5]. Dalam aplikasinya, algoritma Naïve Bayes dan *Support Vector Machine* (SVM) banyak digunakan karena dinilai efektif dalam klasifikasi teks. Naïve Bayes unggul dalam kecepatan komputasi dan efisiensi dalam

memproses data teks, sedangkan *Support Vector Machine* (SVM) dikenal dapat mencapai akurasi tinggi pada data berdimensi tinggi yang dihasilkan dari pembobotan fitur TF-IDF (*Term Frequency-Inverse Document Frequency*) maupun *Bag of Words* [6]. Representasi fitur teks tersebut dipilih karena dapat memberikan bobot informatif pada kata-kata penting dalam dokumen, serta kompatibel dengan kedua algoritma tersebut.

Beberapa penelitian telah menerapkan algoritma Naïve Bayes maupun Support Vector Machine (SVM) untuk analisis sentimen pada domain skincare dan produk kecantikan. Astuti dan Astuti [3] menggunakan Naïve Bayes berbasis Particle Swarm Optimization (PSO) untuk analisis sentimen ulasan skincare dan memperoleh peningkatan performa klasifikasi. Rahmansyah [7] juga menunjukkan bahwa Multinomial Naïve Bayes mampu mengklasifikasikan ulasan produk skincare pada platform Sociolla dengan akurasi yang cukup baik. Sementara itu, penelitian Rahmadzani [8] menunjukkan bahwa SVM memperoleh akurasi lebih tinggi dibandingkan Naïve Bayes dalam analisis sentimen ulasan skincare pada platform Female Daily.

Hasil serupa juga ditemukan oleh Jannah [9] yang membandingkan Naïve Bayes dan SVM pada ulasan produk beauty care di marketplace Shopee. Penelitian tersebut menunjukkan bahwa SVM menghasilkan rata-rata akurasi yang lebih tinggi dibandingkan Naïve Bayes ketika menggunakan representasi TF-IDF dan validasi silang. Selain itu, Setiawan [10] melaporkan bahwa SVM menunjukkan performa yang lebih baik dibandingkan Naïve Bayes pada analisis sentimen komentar TikTok terkait produk skincare, terutama setelah dilakukan penanganan ketidakseimbangan data menggunakan SMOTE.

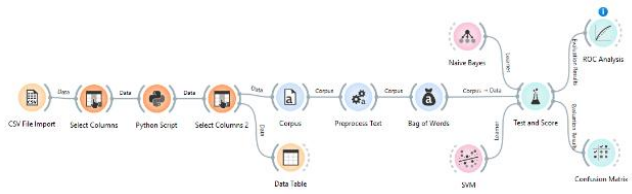
Meskipun berbagai penelitian telah membandingkan performa Naïve Bayes dan SVM, sebagian besar penelitian menggunakan dataset, teknik prapemrosesan, representasi fitur, serta metode evaluasi yang berbeda-beda sehingga hasil yang diperoleh sulit dibandingkan secara langsung. Selain itu, sebagian besar penelitian masih berfokus pada metrik evaluasi konvensional seperti accuracy, precision, recall, dan F1-score, sementara pengaruh ketidakseimbangan kelas terhadap performa model belum banyak dibahas secara mendalam.

Berbeda dengan ulasan e-commerce umum, ulasan produk skincare memiliki karakteristik yang lebih kompleks karena dipengaruhi oleh kondisi kulit individu, pengalaman penggunaan jangka panjang, serta berbagai aspek produk seperti tekstur, kandungan aktif, tingkat hidrasi, dan potensi efek samping. Selain itu, ulasan skincare sering menggunakan istilah khusus kecantikan dan bahasa informal yang dapat meningkatkan kompleksitas proses klasifikasi sentimen.

Berdasarkan kesenjangan tersebut, penelitian ini menawarkan tiga kontribusi utama. Pertama, perbandingan dilakukan menggunakan pipeline eksperimen yang seragam sehingga perbedaan hasil lebih mencerminkan karakteristik algoritma yang digunakan. Kedua, evaluasi model tidak hanya menggunakan metrik konvensional tetapi juga menambahkan Matthews Correlation Coefficient (MCC) dan Area Under Curve (AUC) untuk memberikan gambaran yang lebih representatif pada kondisi data tidak seimbang. Ketiga, penelitian ini menggunakan dataset Sephora Products and Skincare Reviews yang terdiri atas 602.130 ulasan sehingga dapat memberikan gambaran empiris mengenai perilaku Naïve Bayes dan SVM pada dataset ulasan skincare berskala besar dengan distribusi kelas yang tidak seimbang.

II. METODE

Penelitian ini menggunakan desain kuantitatif dengan pendekatan komputasional-eksperimental untuk membandingkan performa dua algoritma klasifikasi teks secara objektif dan terukur. Seluruh proses eksperimen dilakukan menggunakan perangkat lunak Orange versi 3.40.0, yang menyediakan antarmuka visual berbasis alur kerja (*workflow*) untuk membangun pipa analisis data yang modular [11]. Alur penelitian terdiri dari tujuh tahapan berurutan, yaitu pengumpulan dataset, inspeksi dan seleksi data, pelabelan sentimen, pra-pemrosesan teks, ekstraksi fitur, pengembangan model klasifikasi, dan evaluasi performa model. Setiap tahapan dirancang untuk saling mendukung sehingga perbandingan antara kedua algoritma dapat dilakukan di bawah kondisi eksperimen yang sepenuhnya seragam. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian menggunakan Orange Data Mining

A. Pengumpulan Data

Tahap pertama adalah pengumpulan dataset. Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan judul *Sephora Products & Skincare Reviews* yang disusun oleh Nady Inky pada tahun 2023. Dataset ini dipilih karena memuat kumpulan ulasan produk *skincare* dalam jumlah besar dengan atribut yang relevan untuk analisis sentimen, seperti teks ulasan, nama produk, rating, jenis kulit pengguna, serta metadata pendukung lainnya. Dataset diunduh dalam format CSV, kemudian dipilih salah satu berkas yaitu review_0-250 untuk memudahkan proses analisis data.

B. Inspeksi dan Seleksi Data

Tahap kedua adalah inspeksi dan seleksi data awal. Widget Data Table digunakan untuk menganalisis dataset yang telah diimpor ke Orange Data Mining. Hasil pemeriksaan menunjukkan bahwa dataset terdiri dari 602.130 entri dengan 14 fitur dan 5 atribut metadata. Tahap ini

bertujuan memastikan seluruh data telah dimuat dengan benar dan siap digunakan pada proses analisis berikutnya.

Selanjutnya dilakukan seleksi atribut menggunakan widget Select Columns. Pada tahap ini dipilih atribut yang relevan untuk proses analisis sentimen, yaitu atribut rating dan skin_type sebagai fitur, sedangkan *review_text, product_id, product_name,* dan *brand_name* digunakan sebagai metadata. Atribut lain yang tidak memiliki relevansi langsung terhadap proses klasifikasi sentimen dikeluarkan dari analisis untuk mengurangi kompleksitas data dan meningkatkan efisiensi komputasi. Karakteristik dataset ditunjukkan pada Tabel 1.

TABEL I
KARAKTERISTIK DATASET

Komponen	Deskripsi
Sumber	Kaggle
Judul	Sephora Products & Skincare Reviews
Penyusun	Nady Inky (2023)
Format	Csv
Jumlah Data	602.130 Entri
Jumlah Fitur	14 Fitur
Jumlah Meta Atribut	5 Meta Atribut
Bahasa	Bahasa Inggris
Field Utama	Review_Text, Product_Name, Rating, Skin_Type
Metadata Pendukung	Product_Id, Brand_Name, Review_Title, Author_Id
Tools Analisis	Orange Data Mining Versi 3.40.0
Metode Ekstraksi Fitur	Bag Of Words (Bow)
Algoritma Klasifikasi	Naïve Bayes Dan Support Vector Machine (Svm)
Metode Evaluasi	Accuracy, Precision, Recall, F1-Score, Auc, Dan Mcc

C. Pelabelan Sentimen

Tahap ketiga adalah pelabelan sentimen menggunakan widget Python Script. Dataset asli tidak menyediakan label sentimen sehingga dilakukan pelabelan otomatis berdasarkan nilai rating yang diberikan pengguna. Pendekatan ini dipilih karena rating dapat digunakan sebagai indikator tingkat kepuasan pengguna terhadap produk yang diulas dan banyak dimanfaatkan dalam penelitian analisis sentimen untuk membentuk dataset berlabel secara otomatis[12].

TABEL II
KATEGORI SENTIMEN BERDASARKAN NILAI RATING

Rating	Label Sentimen
1 – 2	Negatif
3	Netral
4 – 5	Positif

Atribut baru bernama *sentiment* dibentuk berdasarkan aturan pelabelan yang ditunjukkan pada Tabel II. Proses ini menghasilkan tiga kategori sentimen, yaitu positif, netral, dan negatif, yang selanjutnya digunakan sebagai variabel target dalam proses klasifikasi. Pembagian tersebut bertujuan untuk memetakan opini pengguna sesuai dengan tingkat kepuasan yang direpresentasikan oleh rating produk.

Hasil pelabelan menunjukkan distribusi sentimen yang tidak seimbang, dimana kelas positif mendominasi sebesar 82,37%, sedangkan kelas netral dan negatif masing-masing sebesar 7,31% dan 10,32. Kondisi ini menunjukkan adanya *class imbalance* yang berpotensi memengaruhi performa model klasifikasi, terutama dalam mengenali kelas minoritas. Tabel III menunjukan distribusi kelas sentimen hasil pelabelan.

TABEL III
DISTRIBUSI KELAS SENTIMEN HASIL PELABELAN

Sentimen	Jumlah Data	Presentase
Positif	495.956	82,37%
Netral	44.018	7,31%
Negatif	62.156	10,32%
Total	602.130	100%

Meskipun pendekatan ini memungkinkan proses pelabelan dilakukan secara otomatis pada dataset berukuran besar, terdapat keterbatasan yang perlu diperhatikan. Rating numerik tidak selalu sepenuhnya merepresentasikan isi teks ulasan. Dalam beberapa kasus, pengguna dapat memberikan rating tinggi tetapi tetap menyampaikan kritik terhadap aspek tertentu dari produk, seperti aroma, kemasan, atau harga. Sebaliknya, pengguna juga dapat memberikan rating rendah meskipun terdapat beberapa aspek produk yang dinilai positif. Oleh karena itu, hasil penelitian perlu diinterpretasikan dengan mempertimbangkan potensi ketidaksesuaian antara rating dan isi teks ulasan.

Setelah label sentimen terbentuk, dilakukan seleksi atribut kembali menggunakan widget Select Columns kedua. Pada tahap ini atribut *sentiment* ditetapkan sebagai variabel target, atribut *skin_type* digunakan sebagai fitur tambahan, sedangkan atribut *rating* dihapus dari fitur untuk menghindari kebocoran data (*data leakage*) karena label sentimen dibentuk langsung dari atribut tersebut. Atribut *review_text, product_id, product_name,* dan *brand_name* dipertahankan sebagai metadata. Langkah ini dilakukan untuk memastikan bahwa proses klasifikasi benar-benar didasarkan pada informasi yang terkandung dalam teks ulasan dan bukan pada nilai rating yang digunakan dalam proses pelabelan.

D. Prapemrosesan Teks

Tahap keempat adalah prapemrosesan teks menggunakan widget Corpus dan Preprocess Text yang bertujuan mengurangi noise serta menyeragamkan data sebelum dilakukan ekstraksi fitur [13].

Atribut review_text ditetapkan sebagai sumber teks utama, sedangkan product_name digunakan sebagai judul dokumen. Bahasa korpus diatur ke Bahasa Inggris sesuai dengan bahasa yang digunakan pada seluruh ulasan. Tahapan

prapemrosesan yang diterapkan meliputi konversi huruf menjadi huruf kecil (lowercase), penghapusan URL, tokenisasi menggunakan metode Word and Punctuation, penghapusan stopwords bahasa Inggris, penghapusan angka, serta penyaringan frekuensi dokumen dengan rentang relatif 0,01 hingga 0,90. Tahap ini penting karena data ulasan pengguna sering kali mengandung simbol, tautan, angka, dan variasi penulisan yang dapat menurunkan performa model klasifikasi.

E. Ekstraksi Fitur

Tahap kelima adalah ekstraksi fitur. Teks yang telah melalui proses prapemrosesan diubah menjadi representasi numerik menggunakan metode Bag of Words (BoW), dimana setiap dokumen direpresentasikan sebagai vektor berdasarkan frekuensi kemunculan kata dalam dokumen. Metode Bag of Words banyak digunakan dalam klasifikasi teks karena mampu mengubah data teks tidak terstruktur menjadi fitur numerik yang dapat diproses oleh algoritma machine learning serta memiliki implementasi yang sederhana dan efisien secara komputasi[14],[15].

Pada penelitian ini, representasi fitur dilakukan menggunakan widget Bag of Words pada Orange Data Mining. Hasil transformasi menghasilkan matriks fitur berdimensi tinggi yang kemudian digunakan sebagai masukan bagi algoritma Naïve Bayes dan Support Vector Machine (SVM) pada tahap klasifikasi.

Penggunaan representasi fitur yang sama pada kedua algoritma bertujuan untuk menjaga konsistensi eksperimen sehingga perbedaan performa yang diperoleh dapat diatribusikan pada karakteristik algoritma klasifikasi yang dibandingkan, bukan akibat perbedaan metode representasi fitur yang digunakan.

Meskipun Bag of Words memiliki keunggulan dalam kesederhanaan dan efisiensi komputasi, metode ini tidak mempertimbangkan urutan maupun hubungan semantik antar kata. Oleh karena itu, penggunaan representasi fitur lain seperti TF-IDF, Word2Vec, FastText, atau transformer embeddings dapat menjadi pengembangan pada penelitian selanjutnya.

F. Model Klasifikasi

Tahap keenam adalah pembangunan model klasifikasi menggunakan dua algoritma pembelajaran mesin, yaitu Naïve Bayes dan Support Vector Machine (SVM). Kedua algoritma dipilih karena merupakan metode yang banyak digunakan dalam penelitian analisis sentimen berbasis teks dan memiliki karakteristik yang berbeda dalam proses pembelajaran data.

Algoritma Naïve Bayes yang digunakan merupakan varian Multinomial Naïve Bayes yang mengacu pada Teorema Bayes [3]. Algoritma ini dipilih karena memiliki kompleksitas komputasi yang rendah, mudah diimplementasikan, serta mampu menghasilkan performa yang baik pada tugas klasifikasi teks dengan jumlah fitur yang besar. Selain itu, Naïve Bayes dikenal efektif dalam menangani data berdimensi tinggi yang dihasilkan dari representasi teks seperti Bag of Words [5].

Sementara itu, algoritma Support Vector Machine (SVM) digunakan sebagai metode pembanding karena memiliki kemampuan dalam membangun *hyperplane* optimal untuk memisahkan kelas pada ruang fitur berdimensi tinggi. Pada penelitian ini, SVM menggunakan kernel linear (LinearSVC) dengan nilai Cost (C) sebesar 10, numerical tolerance sebesar 0,001, dan iteration limit sebanyak 100 iterasi. Kernel linear dipilih karena sesuai dengan karakteristik data teks hasil representasi Bag of Words yang bersifat berdimensi tinggi dan *sparse*, serta memiliki efisiensi komputasi yang lebih baik dibandingkan kernel nonlinier seperti Radial Basis Function (RBF) maupun Polynomial pada kasus klasifikasi teks [1], [16]

Penelitian ini tidak melakukan proses optimasi parameter (*hyperparameter tuning*) maupun penerapan teknik penyeimbangan data seperti SMOTE atau *class weighting.* Tujuannya adalah untuk mengevaluasi performa dasar (*baseline performance*) dari kedua algoritma pada distribusi sentimen asli hasil pelabelan. Dengan pendekatan ini, perbedaan performa yang diperoleh dapat diinterpretasikan sebagai representasi kemampuan masing-masing algoritma dalam menangani karakteristik data yang digunakan.

G. Evaluasi Model

Tahap ketujuh adalah evaluasi kinerja model menggunakan widget Test and Score. Evaluasi dilakukan dengan metode Random Sampling sebanyak dua kali pengulangan (repeat train/test = 2), dengan pembagian data sebesar 80% sebagai data pelatihan dan 20% sebagai data pengujian (training set size = 80%). Opsi *stratified sampling* diaktifkan untuk mempertahankan proporsi kelas sentimen secara konsisten pada data pelatihan dan data pengujian sehingga distribusi kelas tetap terjaga selama proses evaluasi.

Metode random sampling dipilih karena ukuran dataset yang sangat besar, yaitu 602.130 data, sehingga 20% data uji telah menghasilkan sekitar 120.000 data pengujian yang representatif. Pengulangan sebanyak dua kali dilakukan untuk memperoleh hasil evaluasi yang lebih stabil.

Evaluasi model dilakukan menggunakan enam metrik, yaitu *AUC, Accuracy, F1-Score, Precision, Recall, dan Matthews Correlation Coefficient (MCC).* Penggunaan berbagai metrik ini bertujuan memberikan penilaian yang lebih komprehensif terhadap kemampuan klasifikasi masing-masing algoritma, terutama pada dataset yang tidak seimbang (imbalanced dataset). Akurasi saja tidak cukup untuk menggambarkan performa model secara menyeluruh karena model yang hanya memprediksi kelas mayoritas juga dapat memperoleh nilai akurasi yang tinggi [17]

Precision digunakan untuk mengukur tingkat ketepatan prediksi model terhadap suatu kelas, sedangkan *Recall* digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar pada kelas tersebut. *F1-Score* digunakan karena mampu menyeimbangkan nilai *Precision* dan *Recall* sehingga memberikan gambaran performa yang lebih representatif, terutama pada kasus klasifikasi dengan distribusi kelas yang tidak seimbang [17]

AUC (Area Under Curve) digunakan untuk mengukur kemampuan model dalam membedakan antar kelas secara keseluruhan tanpa bergantung pada nilai ambang (*threshold*) tertentu. Nilai AUC yang semakin mendekati 1 menunjukkan kemampuan diskriminasi model yang semakin baik, sedangkan nilai yang mendekati 0,5 mengindikasikan bahwa kemampuan model hampir setara dengan tebakan acak (*random guessing*).

Sementara itu, Matthews Correlation Coefficient (MCC) dipilih karena mampu memberikan penilaian yang lebih adil dan komprehensif pada kondisi ketidakseimbangan kelas yang ekstrem. MCC mempertimbangkan seluruh elemen *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), sehingga mampu mengevaluasi kualitas prediksi model secara lebih menyeluruh dibandingkan akurasi. Beberapa penelitian menunjukkan bahwa MCC merupakan salah satu metrik yang lebih stabil dan representatif untuk mengevaluasi model klasifikasi pada dataset yang tidak seimbang [17].

Untuk memvisualisasikan kemampuan diskriminasi masing-masing model, hasil evaluasi juga ditampilkan dalam bentuk kurva Receiver Operating Characteristic (ROC) menggunakan widget ROC Analysis pada Orange Data Mining. Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai ambang keputusan (*threshold*), sehingga dapat digunakan untuk membandingkan kemampuan klasifikasi kedua algoritma pada setiap kelas sentimen.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil klasifikasi sentimen pada ulasan produk *skincare* Sephora beserta interpretasi analisisnya. Seluruh evaluasi dilakukan pada data uji yang diperoleh dari pembagian data menggunakan metode *random sampling* dengan rasio 80:20. Fitur teks direpresentasikan menggunakan metode *Bag of Words (BoW)* dengan *CountVectorizer*. Representasi ini menghasilkan vektor berdimensi tinggi yang sesuai untuk algoritma klasifikasi teks klasik seperti Naïve Bayes dan SVM, serta memungkinkan perbandingan yang adil karena kedua model menggunakan kondisi eksperimen yang sama.

1. Distribusi Kelas Sentimen

Sebelum membandingkan kinerja model klasifikasi, distribusi kelas sentimen pada dataset perlu diperhatikan karena dapat memengaruhi hasil evaluasi model. Berdasarkan hasil pelabelan sentimen yang telah dijelaskan pada tahap metodologi, distribusi data menunjukkan bahwa kelas sentimen positif mendominasi dataset dengan proporsi sebesar 82,37%, sedangkan sentimen negatif dan netral masing-masing sebesar 10,32% dan 7,31%.

Dominasi sentimen positif menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang baik terhadap produk skincare yang diulas. Pola ini umum ditemukan pada dataset ulasan produk e-commerce karena pengguna yang merasa puas cenderung lebih aktif memberikan ulasan dan rating tinggi dibandingkan pengguna yang tidak puas. Selain itu, produk skincare umumnya digunakan dalam jangka waktu tertentu sehingga pengguna yang telah merasakan manfaat produk lebih terdorong untuk memberikan ulasan positif.

Distribusi kelas tersebut juga menunjukkan adanya kondisi ketidakseimbangan kelas (*class imbalance*) yang cukup tinggi. Rasio kelas positif terhadap kelas netral mencapai lebih dari 11:1, sedangkan rasio kelas positif terhadap kelas negatif mencapai hampir 8:1. Ketimpangan ini berpotensi menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas sehingga nilai akurasi yang tinggi belum tentu mencerminkan kemampuan model dalam mengenali seluruh kelas sentimen secara baik.

Pada kondisi dataset yang tidak seimbang, model dapat memperoleh nilai akurasi yang tinggi hanya dengan memprediksi kelas mayoritas. Oleh karena itu, evaluasi pada penelitian ini tidak hanya berfokus pada nilai akurasi, tetapi juga mempertimbangkan metrik Precision, Recall, F1-Score, Area Under Curve (AUC), dan Matthews Correlation Coefficient (MCC) untuk memberikan gambaran performa model yang lebih komprehensif dan representatif terhadap seluruh kelas sentimen [17].

Kondisi *class imbalance* ini juga menjadi salah satu faktor penting dalam menginterpretasikan hasil perbandingan antara algoritma Naïve Bayes dan Support Vector Machine pada bagian selanjutnya. Performa kedua algoritma tidak hanya dipengaruhi oleh karakteristik metode klasifikasi yang digunakan, tetapi juga oleh kemampuan masing-masing algoritma dalam menangani distribusi data yang sangat didominasi oleh kelas sentimen positif.

2. Hasil Evaluasi Model

Perbandingan kinerja kedua algoritma dilakukan menggunakan dataset, tahapan prapemrosesan, dan representasi fitur Bag of Words yang sama dengan metode random sampling 80:20. Hasil evaluasi secara keseluruhan disajikan pada Tabel 4.

TABEL IV
HASIL EVALUASI MODEL

Model	AUC	CA	F1	Precision	Recall	MCC
Naive Bayes	0,784	0,820	0,764	0,749	0,820	0,153
SVM	0,497	0,823	0,743	0,695	0,823	0,000

Berdasarkan Tabel IV, kedua algoritma memperoleh nilai akurasi di atas 82%, namun menunjukkan perbedaan yang cukup besar pada metrik evaluasi lainnya. Naïve Bayes unggul pada metrik AUC (0,784 dibandingkan 0,497), F1-Score (0,764 dibandingkan 0,743), Precision (0,749 dibandingkan 0,695), dan MCC (0,153 dibandingkan 0,000). Sementara itu, SVM hanya sedikit lebih unggul pada metrik Accuracy (0,823 dibandingkan 0,820) dan Recall (0,823 dibandingkan 0,820).

Perbedaan hasil tersebut menunjukkan bahwa nilai akurasi yang tinggi tidak selalu mencerminkan kemampuan model dalam mengenali seluruh kelas sentimen, terutama pada dataset yang memiliki distribusi kelas tidak seimbang. Pada kondisi *class imbalance*, model dapat memperoleh nilai akurasi tinggi hanya dengan memprediksi kelas mayoritas secara dominan sehingga performa pada kelas minoritas tidak tercermin secara memadai melalui metrik akurasi saja [17].

Aspek yang paling penting adalah nilai MCC pada SVM yang mencapai 0,000. Nilai tersebut menunjukkan bahwa model SVM tidak mampu membangun hubungan yang bermakna antara hasil prediksi dan label sebenarnya. MCC yang mendekati nol mengindikasikan bahwa performa model secara keseluruhan hampir setara dengan tebakan acak, meskipun nilai akurasinya mencapai 82,3% [17]. Kondisi ini terjadi karena sebagian besar data diprediksi sebagai kelas mayoritas, yaitu sentimen positif, sehingga model memperoleh akurasi tinggi tetapi tidak mampu mengenali kelas minoritas secara efektif. Fenomena ini umum terjadi pada dataset yang mengalami ketidakseimbangan kelas, dimana model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas [18].

Interpretasi tersebut diperkuat oleh nilai AUC SVM sebesar 0,497 yang sangat dekat dengan nilai 0,5 yang merepresentasikan kemampuan diskriminasi setara dengan klasifikasi acak. Sebaliknya, Naïve Bayes memperoleh nilai AUC sebesar 0,784 yang menunjukkan kemampuan diskriminasi yang jauh lebih baik dalam membedakan kelas sentimen positif, netral, dan negatif.

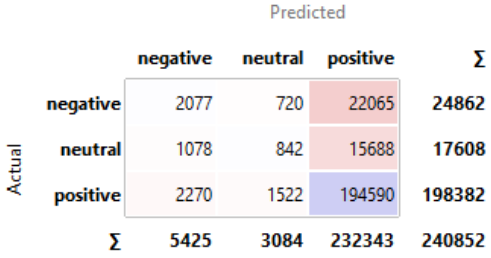
Hasil ini mengindikasikan bahwa pada kondisi dataset yang sangat tidak seimbang, Naïve Bayes mampu mempertahankan kemampuan klasifikasi yang lebih baik dibandingkan SVM pada konfigurasi parameter yang digunakan dalam penelitian ini. Temuan ini sejalan dengan penelitian Jasmarizal [1] yang menunjukkan bahwa performa SVM dapat menurun pada kondisi ketidakseimbangan kelas yang tinggi apabila tidak disertai teknik penyeimbangan data seperti *Synthetic Minority Oversampling Technique* (SMOTE). Oleh karena itu, keunggulan Naïve Bayes pada penelitian ini perlu dipahami dalam konteks karakteristik dataset dan konfigurasi eksperimen yang digunakan, sehingga tidak dapat digeneralisasikan bahwa Naïve Bayes selalu lebih unggul dibandingkan SVM pada seluruh kasus analisis sentimen.

3. Analisis Confusion Matrix

Analisis confusion matrix dilakukan untuk memberikan gambaran yang lebih rinci mengenai pola kesalahan klasifikasi yang dihasilkan oleh masing-masing model pada setiap kelas sentimen.

TABEL V
CONFUSION MATRIX NAIVE BAYES

Actual/predicted	**Negatif**	**Netral**	**Positif**
Negatif	2.077	720	22.065
Netral	1.078	842	15.688
Positif	2.270	1.522	194.590

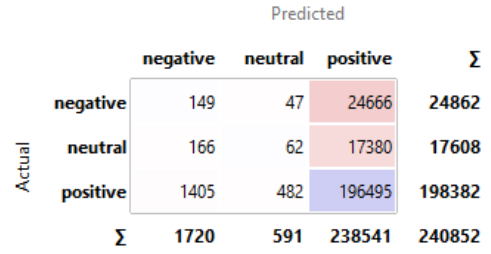


Gambar 2. Confusion Matrix Naive Bayes

Berdasarkan Tabel V dan Gambar 2, Naïve Bayes berhasil mengklasifikasikan kelas positif dengan sangat baik, yaitu sebanyak 194.590 dari 198.382 data (98,09%) berhasil diprediksi dengan benar. Namun, performa model menurun pada kelas minoritas. Pada kelas negatif, hanya 2.077 dari 24.862 data (8,35%) yang berhasil diklasifikasikan dengan benar, sedangkan pada kelas netral hanya 842 dari 17.608 data (4,78%) yang berhasil diprediksi secara tepat. Kesalahan terbesar terjadi ketika data negatif dan netral diprediksi sebagai sentimen positif. Meskipun demikian, Naïve Bayes masih menunjukkan kemampuan mengenali pola leksikal pada kelas minoritas yang ditunjukkan oleh adanya sejumlah data negatif dan netral yang berhasil diklasifikasikan dengan benar.

TABEL VI
CONFUSION MATRIX SVM

Actual/predicted	Negatif	Netral	Positif
Negatif	149	47	24.666
Netral	166	62	17.380
Positif	1.405	482	196.495



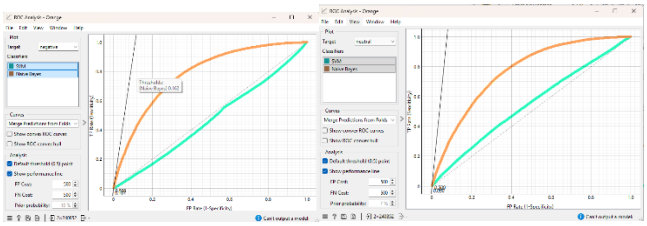
Gambar 3. Confusion Matrix SVM

Berdasarkan Tabel VI dan Gambar 3, pola kesalahan klasifikasi pada SVM jauh lebih buruk dibandingkan Naïve Bayes. Hanya 149 dari 24.862 data negatif (0,60%) dan 62 dari 17.608 data netral (0,35%) yang berhasil diprediksi dengan benar. Hampir seluruh data dari semua kelas diprediksi sebagai sentimen positif, termasuk 24.666 data negatif dan 17.380 data netral yang salah diklasifikasikan sebagai positif. Kondisi ini menjelaskan mengapa nilai MCC SVM mendekati nol, yaitu karena model tidak mampu membedakan kelas minoritas secara efektif dan hanya mengandalkan prediksi kelas mayoritas untuk memperoleh akurasi yang tinggi.

Perbedaan pola klasifikasi kedua model menunjukkan bahwa Naïve Bayes lebih robust dalam menangani ketidakseimbangan kelas dibandingkan SVM dengan kernel linear pada konfigurasi yang digunakan. Pendekatan probabilistik pada Naïve Bayes masih mampu menangkap sebagian pola leksikal kelas minoritas, sedangkan SVM cenderung meminimalkan kesalahan global dengan memprediksi kelas mayoritas pada hampir seluruh data.

4. Analisis Kurva ROC

Kurva *ROC (Receiver Operating Characteristic)* digunakan untuk memberikan representasi visual terhadap kemampuan diskriminasi masing-masing model. Kurva ROC menggambarkan hubungan antara *True Positive Rate* (sensitivitas) dan *False Positive Rate* (1-spesifisitas) pada berbagai nilai *threshold*. Model yang lebih baik akan menghasilkan kurva yang semakin mendekati sudut kiri atas grafik. Analisis ROC difokuskan pada kelas negatif dan netral karena kedua kelas tersebut merupakan kelas minoritas yang lebih informatif dalam menilai kemampuan model pada kondisi ketidakseimbangan kelas yang ekstrem.



Gambar 4. ROC Curve Negatif dan Netral

Berdasarkan Gambar 4, terdapat perbedaan yang sangat jelas antara kurva ROC Naïve Bayes dan SVM pada kelas negatif maupun netral. Kurva Naïve Bayes berada lebih tinggi dan lebih dekat ke sudut kiri atas dibandingkan kurva SVM yang hampir mengikuti garis diagonal. Garis diagonal pada grafik ROC merepresentasikan kemampuan diskriminasi yang setara dengan klasifikasi acak (AUC = 0,5). Oleh karena itu, posisi kurva SVM yang mendekati garis tersebut secara visual menegaskan bahwa model SVM tidak mampu membedakan kelas minoritas dari kelas mayoritas secara efektif. Temuan ini sejalan dengan nilai AUC SVM sebesar 0,497 dan semakin memperkuat kesimpulan bahwa Naïve Bayes memiliki kemampuan diskriminasi yang lebih baik pada dataset dengan ketidakseimbangan kelas yang tinggi.

5. Pembahasan dan Implikasi

Secara keseluruhan, hasil penelitian menunjukkan bahwa Naïve Bayes memberikan performa yang lebih seimbang dan lebih andal dibandingkan SVM dalam klasifikasi sentimen ulasan produk skincare pada dataset yang sangat tidak seimbang. Keunggulan Naïve Bayes pada metrik AUC, F1-Score, Precision, dan MCC menunjukkan bahwa algoritma ini lebih efektif dalam menangani distribusi kelas yang tidak seimbang tanpa memerlukan penyesuaian tambahan.

Kegagalan SVM pada penelitian ini sejalan dengan karakteristik yang telah dilaporkan dalam berbagai literatur, yaitu bahwa SVM dengan kernel linear dan parameter standar sangat rentan terhadap ketidakseimbangan kelas yang ekstrem karena fungsi keputusan model cenderung terdorong ke arah kelas mayoritas Kondisi ini berbeda dengan beberapa penelitian sebelumnya yang melaporkan bahwa SVM mampu mengungguli Naïve Bayes pada tugas analisis sentimen dengan distribusi kelas yang lebih seimbang. Hal tersebut menunjukkan bahwa keunggulan suatu algoritma sangat dipengaruhi oleh karakteristik dataset yang digunakan, terutama tingkat ketidakseimbangan kelas dan domain teks yang dianalisis.

Dari sisi praktis, dominasi sentimen positif sebesar 82,37% menunjukkan tingginya tingkat kepuasan pengguna terhadap produk *skincare* Sephora. Temuan ini dapat dimanfaatkan oleh perusahaan kosmetik untuk memahami persepsi konsumen secara lebih mendalam. Ulasan negatif dan netral yang jumlahnya lebih sedikit tetap memiliki nilai penting karena dapat memberikan masukan terkait aspek produk yang masih perlu ditingkatkan. Dengan mengidentifikasi pola kata yang muncul pada ulasan negatif dan netral, perusahaan dapat memperoleh wawasan yang lebih spesifik untuk mendukung pengembangan produk di masa mendatang.

Selain itu, penelitian ini menegaskan pentingnya penggunaan berbagai metrik evaluasi pada dataset yang tidak seimbang. Hasil penelitian menunjukkan bahwa akurasi saja dapat memberikan kesan yang menyesatkan. SVM memperoleh akurasi sebesar 82,3%, namun memiliki nilai MCC sebesar 0,000 yang menunjukkan bahwa model tidak mempelajari pola data secara bermakna. Oleh karena itu, penggunaan MCC, AUC, dan F1-Score secara bersamaan mampu memberikan gambaran performa model yang lebih jujur dan komprehensif dibandingkan hanya mengandalkan akurasi

6. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam menginterpretasikan hasil penelitian. Pertama, pelabelan sentimen dilakukan secara otomatis berdasarkan nilai rating, sehingga berpotensi menghasilkan noise apabila terdapat ketidaksesuaian antara nilai rating dan isi ulasan. Sebagai contoh, pengguna dapat memberikan nilai rating yang tinggi, tetapi isi ulasan mengandung kritik atau ketidakpuasan terhadap produk.

Kedua, distribusi kelas yang sangat tidak seimbang, dengan dominasi sentimen positif sebesar 82,37%, memengaruhi kemampuan model dalam mengenali kelas minoritas secara optimal. Kondisi ini terutama terlihat pada algoritma Support Vector Machine (SVM) yang menunjukkan kesulitan dalam membedakan kelas negatif dan netral dari kelas mayoritas.

Ketiga, representasi fitur menggunakan metode Bag of Words (BoW) belum mampu menangkap konteks semantik, sarkasme, maupun makna implisit yang sering ditemukan dalam teks ulasan pengguna. Akibatnya, informasi

kontekstual yang penting dalam menentukan sentimen dapat terabaikan selama proses klasifikasi.

Keempat, konfigurasi parameter SVM yang digunakan masih berupa pengaturan standar tanpa penerapan teknik penyeimbangan data (data balancing) maupun optimasi parameter yang lebih mendalam. Oleh karena itu, performa optimal algoritma SVM belum sepenuhnya dievaluasi dalam penelitian ini.

Kelima, penelitian ini belum mengevaluasi aspek efisiensi komputasi seperti waktu pelatihan (training time), waktu inferensi (inference time), dan penggunaan memori pada masing-masing algoritma. Oleh karena itu, perbandingan yang dilakukan masih berfokus pada performa klasifikasi dan belum mencakup aspek efisiensi implementasi praktis yang juga penting dalam penerapan model pada dataset berskala besar.

Keenam, penelitian ini belum melakukan uji signifikansi statistik terhadap perbedaan performa kedua model sehingga hasil yang diperoleh belum dapat disimpulkan sebagai perbedaan yang signifikan secara statistik.

Keterbatasan tersebut membuka peluang bagi penelitian selanjutnya untuk menerapkan teknik penyeimbangan data, seperti Synthetic Minority Oversampling Technique (SMOTE) atau class weighting, menggunakan representasi fitur yang lebih kaya seperti TF-IDF, word embedding, maupun model berbasis transformer, serta melakukan anotasi data secara manual atau semiotomatis guna meningkatkan kualitas data pelatihan dan performa model klasifikasi sentimen. Selain itu, penelitian selanjutnya dapat menambahkan baseline lain seperti Logistic Regression, Random Forest, XGBoost, maupun model deep learning untuk memperoleh perbandingan yang lebih komprehensif.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Multinomial Naïve Bayes dan Support Vector Machine (SVM) dalam klasifikasi sentimen ulasan produk *skincare* Sephora menggunakan dataset *Sephora Products and Skincare Reviews*. Hasil penelitian menunjukkan bahwa distribusi sentimen pada dataset didominasi oleh sentimen positif sebesar 82,37%, diikuti sentimen negatif sebesar 10,32% dan sentimen netral sebesar 7,31%, yang menunjukkan adanya ketidakseimbangan kelas yang cukup tinggi dan berpotensi memengaruhi performa model klasifikasi.

Berdasarkan hasil evaluasi, algoritma Naïve Bayes menunjukkan performa yang lebih baik dibandingkan SVM pada sebagian besar metrik evaluasi. Naïve Bayes memperoleh nilai AUC sebesar 0,784, F1-Score sebesar 0,764, Precision sebesar 0,749, dan MCC sebesar 0,153, sedangkan SVM memperoleh nilai AUC sebesar 0,497, F1-Score sebesar 0,743, Precision sebesar 0,695, dan MCC sebesar 0,000. Meskipun SVM memiliki nilai Accuracy dan Recall yang sedikit lebih tinggi, nilai MCC yang mendekati nol menunjukkan bahwa model belum mampu melakukan klasifikasi secara efektif pada kelas minoritas.

Analisis *confusion matrix* dan kurva ROC memperkuat temuan bahwa Naïve Bayes memiliki kemampuan diskriminasi yang lebih baik dalam membedakan kelas sentimen negatif, netral, dan positif dibandingkan SVM pada kondisi dataset yang sangat tidak seimbang. Hasil penelitian ini menunjukkan bahwa Naïve Bayes memberikan performa yang lebih baik dibandingkan SVM pada konfigurasi eksperimen dan karakteristik dataset yang digunakan dalam penelitian ini. Namun, temuan tersebut tidak dapat digeneralisasikan bahwa Naïve Bayes selalu lebih unggul dibandingkan SVM pada seluruh kasus analisis sentimen karena performa kedua algoritma sangat dipengaruhi oleh karakteristik data, representasi fitur, serta konfigurasi parameter yang digunakan.

Selain itu, penelitian ini menegaskan pentingnya penggunaan berbagai metrik evaluasi seperti AUC, F1-Score, dan MCC dalam menilai performa model klasifikasi, karena akurasi saja tidak cukup untuk menggambarkan kemampuan model secara menyeluruh pada dataset yang mengalami ketidakseimbangan kelas, sehingga diperlukan metrik tambahan yang mampu mengevaluasi performa model secara lebih komprehensif.

Meskipun demikian, hasil penelitian ini perlu diinterpretasikan dengan mempertimbangkan keterbatasan yang telah dijelaskan pada bagian sebelumnya, terutama terkait pelabelan otomatis berbasis rating, ketidakseimbangan distribusi kelas, dan konfigurasi model yang digunakan.

DAFTAR PUSTAKA

[1] Jasmirzal, Rahmaddeni, Junadhi, and M. Khairul Anam, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap Produk Skincare," *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 2024–1438, Feb. 2024.

[2] N. Widhiyanta, I. Muhandhis, R. Syadillal Jannah, and L. Alfina Wulansari, "Analisis Sentimen Ulasan Produk Moisturizer Skintific Di Tokopedia Menggunakan Support Vector Machine," 2025.

[3] T. Astuti and Y. Astuti, "Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO)," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 1806, Oct. 2022, doi: 10.30865/mib.v6i4.4119.

[4] D. A. WP, J. D. Firizqi, and Z. A. Amalia, "Analisis Sentimen Produk Skincare Somethinc Niacinamide di Female Daily dengan Naïve Bayes Classifier," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 946, Apr. 2024, doi: 10.30865/mib.v8i2.7571.

[5] A. F. Setyaningsih, D. Septiyani, and S. R. Widiasari, "Implementasi Algoritma NaÃ¨ve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 1, pp. 224–235, May 2023, doi: 10.37012/jtik.v9i1.1409.

[6] Nicholas and R. Sutomo, "Comparative Analysis of Sentiment Analysis Using the Support Vector Machine and Naive Bayes Algorithm on Cryptocurrencies," 2021. [Online]. Available: www.jmis.site

[7] F. Rahmansyah, S. Lestari, and S. Yusuf Irianto, "Sentiment Analysis of Skincare Products Using Lexicon and Multinomial Naive Bayes on The Sociolla Website," *Architecture and High Performance Computing*, vol. 7, no. 4, 2025, doi: 10.47709/cnahpc.v7i4.7151.

[8] R. F. Rahmadzani, R. W. Pratiwi, and A. N. Paradita, "Comparison of Naive Bayes and Support Vector Machine (SVM) Methods in

Female Daily Skincare Sentiment Analysis," *Radiant*, vol. 6, no. 2, pp. 99–108, Jun. 2025, doi: 10.52187/rdt.v6i2.316.

[9] N. Z. B. Jannah and K. Kusnawi, "Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, pp. 727–733, Mar. 2024, doi: 10.33395/sinkron.v8i2.13559.

[10] T. Setiawan, S. Liem, and D. M. R. Pribadi, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare," 2024. [Online]. Available: https://jurnal.politap.ac.id/index.php/aicoms

[11] A. Fathiarahma, A. Voutama, T. Ridwan, and N. Heryana, "Analisis Text Mining Klasifikasi Kegiatan Keluarga menggunakan Orange dengan Metode Naive Bayes," *Jurnal Teknologi Terpadu*, Dec. 2022.

[12] M. Hamka, N. Alfatari, and D. Ratna Sari, "Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 64, Sep. 2022, doi: 10.30865/json.v4i1.4740.

[13] B. Darmawan, A. Dwi Laksito, M. Resa, A. Yudianto, and A. Sidauruk, "Analisis Perbandingan Ekstraksi Fitur Teks pada Sentimen Analisis Kenaikan Harga BBM," *Krea-TIF: Jurnal Teknik Informatika*, vol. 11, no. 1, pp. 53–63, 2023, doi: 10.32832/krea-tif.v11i1.13819.

[14] C. Ali, "Klasifikasi Topik Berita Politik Menggunakan Model Logistic Regression Dan Fitur Bag Of Words," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 3, pp. 282–291, Sep. 2025, doi: 10.69916/jkbti.v4i3.375.

[15] A. D. M. Putri, N. Sulistianingsih, and R. Rismayati, "Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF terhadap Akurasi Klasifikasi Sentimen Teks Multi-Domain," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 4, pp. 675–688, Oct. 2025, doi: 10.35746/jtim.v7i4.756.

[16] V. D. Yunanda and N. Hendrastuty, "Perbandingan Kernel Polynomial dan RBF Pada Algoritma SVM Untuk Analisis Sentimen Skincare di Indonesia," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 726, Apr. 2024, doi: 10.30865/mib.v8i2.7425.

[17] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01313-4.

[18] W. I. Salsabila and C. B. Vista, "Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan,"
2021.[Online].Available:https://jurnal.pcr.ac.id/index.php/jkt/