



Analisis Sentimen Ulasan Skintific 5x *Ceramide Barrier Repair Moisture Gel* Pada Platform *Female Daily* Menggunakan SVM Dan Naïve Bayes

Yola Tirta Nurfajriya¹, Pinkan Prilia Mahmudah², Hana Aurora Aida Rahma³, Hanif Zaidan Sinaga^{4*}

^{1,2,3,4} rogram Studi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Ibn Khaldun Bogor

¹yolanrfjry@gmail.com, ²priliapinkan@gmail.com, ³haauaira@gmail.com, ^{4*}hanif.zaidan@gmail.com

Abstrak

Perkembangan industri *skincare* di Indonesia mendorong meningkatnya jumlah ulasan konsumen pada platform digital, salah satunya Female Daily. Ulasan tersebut dapat dimanfaatkan untuk mengetahui persepsi pengguna terhadap suatu produk melalui analisis sentimen. Penelitian ini bertujuan menganalisis sentimen ulasan produk Skintific 5X Ceramide Barrier Repair Moisture Gel serta membandingkan performa algoritma Naïve Bayes dan Support Vector Machine (SVM). Data penelitian diperoleh menggunakan teknik *web scraping* pada platform Female Daily sebanyak 1.029 ulasan. Pelabelan sentimen dilakukan secara manual (*ground truth*) oleh dua anotator independen dan divalidasi menggunakan Cohen's Kappa. Tahapan pengolahan data meliputi *preprocessing*, pembobotan TF-IDF, serta klasifikasi menggunakan Naïve Bayes dan SVM dengan metode Stratified 10-Fold Cross Validation. Hasil pelabelan menunjukkan bahwa sentimen positif mendominasi dataset sebesar 85,62%, diikuti sentimen netral 8,94%, dan sentimen negatif 5,44%. Nilai Cohen's Kappa sebesar 0,6638 menunjukkan tingkat kesepakatan yang tinggi antar anotator. Hasil klasifikasi menunjukkan bahwa Naïve Bayes memperoleh *accuracy* 85,62% dan *macro F1-score* 31,48%, sedangkan SVM memperoleh *accuracy* 88,73% dan *macro F1-score* 58,12%. Penerapan SVM Balanced memberikan hasil terbaik dengan *accuracy* 89,12% dan *macro F1-score* 63,11%. Hasil penelitian menunjukkan bahwa SVM Balanced lebih efektif dibandingkan Naïve Bayes dalam mengklasifikasikan sentimen pada dataset yang tidak seimbang.

Kata Kunci: Analisis Sentimen, Female Daily, Naïve Bayes, Support Vector Machine, TF-IDF.

Abstract

The rapid growth of the skincare industry in Indonesia has led to an increasing number of consumer reviews on digital platforms such as Female Daily. These reviews provide valuable information for understanding user perceptions through sentiment analysis. This study aims to analyze the sentiment of reviews on Skintific 5X Ceramide Barrier Repair Moisture Gel and compare the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms. A total of 1,029 reviews were collected from Female Daily using a *web scraping* technique. Sentiment labels were manually assigned (*ground truth*) by two independent annotators and validated using Cohen's Kappa. The data processing stages included *preprocessing*, TF-IDF feature weighting, and sentiment classification using Naïve Bayes and SVM with Stratified 10-Fold Cross Validation. The labeling results showed that positive sentiment dominated the dataset (85.62%), followed by neutral (8.94%) and negative (5.44%) sentiments. A Cohen's Kappa value of 0.6638 indicated substantial agreement between the annotators. The classification results showed that Naïve Bayes achieved an accuracy of 85.62% with a macro F1-score of 31.48%, while SVM achieved an accuracy of 88.73% with a macro F1-score of 58.12%. The SVM Balanced model produced the best performance, achieving an accuracy of 89.12% and a macro F1-score of 63.11%. These findings indicate that SVM Balanced is more effective than Naïve Bayes for sentiment classification on imbalanced skincare review datasets.

Keywords: Sentiment Analysis, Female Daily, Naïve Bayes, Support Vector Machine, Tf-Idf.

1. Pendahuluan

Perkembangan industri *skincare* di Indonesia meningkat seiring kesadaran masyarakat terhadap kesehatan kulit. Kondisi ini membuat konsumen tidak hanya bergantung pada informasi produsen, tetapi juga pada ulasan pengguna di platform digital. Ulasan menjadi sumber penting untuk menilai kualitas produk dan kepuasan pengguna. Salah satu

produk yang banyak dibahas adalah Skintific 5X Ceramide Barrier Repair Moisture Gel yang diformulasikan untuk menjaga dan memperbaiki skin barrier (Amalia et al., 2023; Khotimah & Arianto, 2025).

Female Daily menjadi salah satu platform utama untuk ulasan produk kecantikan di Indonesia. Ulasan yang tersedia berasal dari pengalaman langsung pengguna sehingga dapat menggambarkan persepsi konsumen terhadap produk. Data ini juga bermanfaat bagi konsumen maupun perusahaan dalam memahami kebutuhan pasar (Yutika et al., 2021). Jumlah ulasan yang besar menghasilkan data teks yang sulit dianalisis secara manual, sehingga diperlukan analisis sentimen berbasis text mining dan NLP untuk mengklasifikasikan opini menjadi positif, negatif, dan netral (Amalia et al., 2023).

Berbagai penelitian sebelumnya telah menerapkan metode machine learning dalam analisis sentimen produk skincare. Amalia et al. (2023) menerapkan metode Naïve Bayes untuk menganalisis sentimen pada ulasan produk moisturizer di Female Daily. Selain itu, Khotimah dan Arianto (Khotimah & Arianto, 2025) menggunakan metode Naïve Bayes untuk menganalisis sentimen ulasan produk sunscreen pada platform Female Daily. Hasil dari berbagai penelitian tersebut menunjukkan bahwa Naïve Bayes masih menjadi salah satu algoritma yang banyak digunakan dalam klasifikasi data teks karena memiliki proses komputasi yang sederhana dan efisien.

Selain Naïve Bayes, Support Vector Machine (SVM) banyak digunakan dalam analisis sentimen karena efektif dalam menangani data teks berdimensi tinggi. Penelitian Pratiwi et al. (2021) menunjukkan penerapan SVM pada ulasan produk kecantikan di Female Daily, sementara Harnelia dan Saputra (2024) serta (Widhiyanta et al., 2025) juga menggunakan SVM pada ulasan produk Skintific dan moisturizer dengan pendekatan TF-IDF. Hasil penelitian tersebut menunjukkan bahwa SVM merupakan metode yang efektif untuk klasifikasi teks.

Penelitian yang membandingkan performa Naïve Bayes dan SVM juga telah banyak dilakukan. Rahmadzani et al. (2025) melakukan perbandingan kedua algoritma tersebut pada data ulasan produk kecantikan. Penelitian lain yang dilakukan oleh Clarisha et al. (2025) serta Lestari dan Hutagalung (2025) juga membandingkan penggunaan Naïve Bayes dan SVM dalam analisis sentimen. Hasil penelitian tersebut menunjukkan bahwa kedua algoritma memiliki karakteristik dan performa yang berbeda dalam proses klasifikasi sentimen sehingga menarik untuk dibandingkan pada objek penelitian yang berbeda. Meskipun penelitian mengenai analisis sentimen produk skincare telah banyak dilakukan, masih terdapat beberapa keterbatasan pada penelitian terdahulu. Sebagian besar penelitian menggunakan pelabelan sentimen otomatis berdasarkan rating maupun pendekatan tertentu sehingga validitas label sentimen yang digunakan masih bergantung pada sistem yang diterapkan. Selain itu, sebagian penelitian hanya menggunakan satu algoritma klasifikasi tanpa melakukan perbandingan dengan algoritma lain. Penelitian yang secara khusus menganalisis ulasan **Skintific 5X Ceramide Barrier Repair Moisture Gel** pada platform Female Daily menggunakan **ground truth manual**, validasi **Cohen's Kappa**, serta perbandingan algoritma **Naïve Bayes** dan **SVM** dengan metode **Stratified 10-Fold Cross Validation** masih relatif terbatas.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna terhadap produk Skintific 5X Ceramide Barrier Repair Moisture Gel pada platform Female Daily menggunakan algoritma Naïve Bayes dan SVM. Data penelitian diperoleh melalui proses web scraping sebanyak 1.029 ulasan. Pelabelan sentimen dilakukan secara manual oleh dua anotator independen dan divalidasi menggunakan Cohen's Kappa untuk mengukur tingkat kesepakatan antar anotator. Selanjutnya, data melalui tahapan preprocessing dan pembobotan TF-IDF sebelum dilakukan proses klasifikasi menggunakan Naïve Bayes dan SVM. Evaluasi model dilakukan menggunakan metode Stratified 10-Fold Cross Validation dengan metrik accuracy, precision, recall, F1-score, macro F1-score, dan confusion matrix.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen untuk mengidentifikasi persepsi pengguna terhadap produk Skintific 5X Ceramide Barrier Repair Moisture Gel berdasarkan ulasan pada platform Female Daily. Analisis sentimen dilakukan dengan membandingkan performa algoritma Naïve Bayes dan SVM mengklasifikasikan sentimen ulasan ke dalam kategori positif, negatif, dan netral. Tahapan penelitian meliputi pengumpulan data, pelabelan sentimen, validasi pelabelan, preprocessing, pembobotan TF-IDF, klasifikasi sentimen, serta evaluasi model.



Gambar 1. Alur dan Skema Penelitian

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna produk Skintific 5X Ceramide Barrier Repair Moisture Gel yang diperoleh dari platform Female Daily. Pengumpulan data dilakukan menggunakan teknik web scraping dengan bantuan bahasa pemrograman Python. Hasil proses scraping menghasilkan sebanyak 1.029 ulasan yang kemudian digunakan sebagai dataset penelitian. Dataset yang telah terkumpul selanjutnya digunakan pada proses pelabelan sentimen dan pengolahan data.

2.2. Pelabelan Sentimen (Ground Truth)

Setelah proses pengumpulan data selesai dilakukan, tahap selanjutnya adalah pelabelan sentimen (*ground truth*) terhadap seluruh dataset. Pelabelan dilakukan secara manual pada 1.029 ulasan produk Skintific 5X Ceramide Barrier Repair Moisture Gel yang diperoleh dari platform Female Daily. Proses pelabelan melibatkan dua orang anotator independen yang memberikan label sentimen berdasarkan isi dan konteks dari setiap ulasan.

Kategori sentimen yang digunakan dalam penelitian ini terdiri atas tiga kelas, yaitu positif, negatif, dan netral. Penggunaan tiga kategori sentimen bertujuan untuk memberikan representasi opini yang lebih komprehensif dibandingkan klasifikasi biner, sehingga ulasan yang bersifat objektif atau tidak menunjukkan kecenderungan sentimen yang dominan tetap dapat terakomodasi dalam proses klasifikasi (Melisa Nur Aini et al., 2024; Yutika et al., 2021).

Sentimen positif diberikan pada ulasan yang menunjukkan kepuasan atau pengalaman baik terhadap produk. Contoh ulasan yang termasuk ke dalam kategori positif adalah *“Melembapkan banget dan bikin skin barrier membaik.”* Sentimen negatif diberikan pada ulasan yang menunjukkan ketidakpuasan atau pengalaman buruk terhadap produk. Contoh ulasan yang termasuk ke dalam kategori negatif adalah *“Muncul beruntusan setelah pemakaian beberapa hari.”*

Sementara itu, sentimen netral diberikan pada ulasan yang bersifat informatif, objektif, atau tidak menunjukkan kecenderungan sentimen positif maupun negatif yang dominan. Contoh ulasan netral adalah *“Teksturnya ringan dan cepat meresap, tetapi wanginya cukup kuat.”* serta *“Kulit terasa lembap, tetapi efek mencerahkannya belum terlihat.”* Hasil pelabelan dari kedua anotator kemudian digunakan sebagai dasar dalam proses validasi kesepakatan label sebelum data digunakan pada tahap klasifikasi sentimen.

2.3. Validasi Pelabelan Menggunakan Cohen's Kappa

Setelah proses pelabelan sentimen selesai dilakukan oleh dua anotator independen, tahap selanjutnya adalah mengukur tingkat kesepakatan antar anotator menggunakan metode Cohen's Kappa. Pengukuran ini dilakukan untuk memastikan bahwa label sentimen yang diberikan memiliki konsistensi yang baik dan tidak terjadi secara kebetulan. Cohen's Kappa merupakan salah satu metode yang umum digunakan untuk mengukur *inter-annotator agreement* pada proses pelabelan data teks (Rahmadzani et al., 2025).

$$K = \frac{Po - Pe}{1 - Pe} \quad (1)$$

Dengan K adalah nilai Cohen's kappa, Po adalah proporsi kesepakatan yang diamati (*observed agreement*), dan Pe adalah proporsi kesepakatan yang terjadi secara acak (*expected agreement*).

Semakin tinggi nilai Cohen's Kappa yang diperoleh, maka semakin tinggi pula tingkat kesepakatan antar anotator dalam memberikan label sentimen. Apabila terdapat perbedaan label antara kedua anotator, dilakukan proses diskusi untuk menentukan label akhir (*gold label*) yang digunakan sebagai label sentimen pada dataset penelitian.

2.4. Text Preprocessing

Tahap preprocessing dilakukan untuk membersihkan dan menyiapkan data teks sebelum dilakukan proses klasifikasi.

Tahapan preprocessing yang digunakan meliputi:

- Cleaning:** Cleaning dilakukan dengan menghapus karakter yang tidak diperlukan seperti URL, simbol, angka, tanda baca, emoji, serta karakter khusus lainnya yang tidak memiliki kontribusi terhadap proses analisis sentimen.
- Case Folding:** Case folding merupakan proses mengubah seluruh huruf pada teks menjadi huruf kecil (lowercase) agar tidak terjadi perbedaan representasi kata akibat penggunaan huruf kapital dan huruf kecil (Khoerunnisa et al., 2025).
- Normalisasi:** Normalisasi dilakukan untuk mengubah kata tidak baku, singkatan, bahasa gaul, maupun variasi penulisan kata menjadi bentuk baku sesuai dengan kaidah bahasa Indonesia (Khoerunnisa et al., 2025).

- d) Tokenizing: Tokenizing merupakan proses pemecahan kalimat menjadi unit-unit kata (token) yang akan digunakan pada tahap analisis berikutnya.
- e) Stopword Removal: Stopword removal dilakukan dengan menghapus kata-kata umum yang tidak memiliki makna sentimen yang signifikan, seperti “dan”, “yang”, “di”, “ke”, serta kata umum lainnya.
- f) Stemming: Stemming merupakan proses mengubah kata berimbuhan menjadi bentuk kata dasarnya. Contohnya kata “membantu” menjadi “bantu” dan kata “menggunakan” menjadi “guna”.

2.5. Pembobotan TF-IDF

Data hasil preprocessing kemudian ditransformasikan ke dalam bentuk numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF merupakan metode pembobotan kata yang memberikan nilai berdasarkan frekuensi kemunculan kata dalam dokumen dan tingkat kemunculannya pada keseluruhan korpus. Kata yang memiliki tingkat relevansi tinggi akan memperoleh bobot yang lebih besar dibandingkan kata yang sering muncul namun kurang informatif (Harmandini & L, 2024). Pembobotan TF-IDF dihitung menggunakan Persamaan (2), sedangkan nilai *Inverse Document Frequency* (IDF) dihitung menggunakan Persamaan (3).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (2)$$

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \quad (3)$$

Dengan $TF(t, d)$ adalah frekuensi kemunculan kata t pada dokumen d , $IDF(t)$ adalah bobot inverse document frequency dari kata t , N adalah jumlah seluruh dokumen, dan $df(t)$ adalah dokumen yang mengandung kata t .

Hasil dari proses pembobotan TF-IDF berupa matriks fitur yang merepresentasikan setiap ulasan dalam bentuk numerik. Matriks fitur tersebut kemudian digunakan sebagai masukan (*input*) pada proses klasifikasi sentimen menggunakan algoritma Naïve Bayes dan SVM.

2.6. Klasifikasi Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu. Algoritma ini banyak digunakan dalam analisis sentimen karena memiliki proses komputasi yang sederhana dan efisien dalam mengolah data teks (Amalia et al., 2023).

2.7. Klasifikasi SVM

SVM merupakan algoritma klasifikasi yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam kelas yang berbeda. Algoritma ini efektif digunakan pada data berdimensi tinggi seperti data teks hasil ekstraksi TF-IDF dan banyak diterapkan pada penelitian analisis sentimen produk skincare (Harnelia, 2024; Widhiyanta et al., 2025; Wijayatun Pratiwi et al., 2021).

2.8. Evaluasi Model

Pada penelitian ini, evaluasi model dilakukan menggunakan metode **Stratified 10-Fold Cross Validation**. Metode ini membagi dataset menjadi sepuluh bagian dengan tetap mempertahankan proporsi setiap kelas sentimen pada masing-masing *fold*. Setiap *fold* digunakan secara bergantian sebagai data uji, sedangkan *fold* lainnya digunakan sebagai data latih. Penggunaan Stratified 10-Fold Cross Validation bertujuan untuk menghasilkan evaluasi yang lebih stabil dan mengurangi bias yang dapat terjadi pada pembagian data tunggal (Jannah & Kusnawi, 2024; Rahmadzani et al., 2025).

Kinerja model dievaluasi menggunakan metrik **accuracy**, **precision**, **recall**, **F1-score**, **macro F1-score**, dan **confusion matrix**. Hasil evaluasi dari algoritma Naïve Bayes dan SVM kemudian dibandingkan untuk menentukan metode yang memiliki performa terbaik dalam klasifikasi sentimen ulasan produk Skintific 5X Ceramide Barrier Repair Moisture Gel.

3. Hasil dan Pembahasan

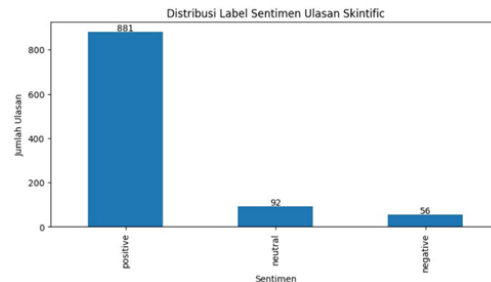
3.1. Pengumpulan Data

Tahap pengumpulan data dilakukan menggunakan teknik *web scraping* pada platform Female Daily dengan objek penelitian berupa produk Skintific 5X Ceramide Barrier Repair Moisture Gel. Platform Female Daily dipilih karena merupakan salah satu komunitas kecantikan terbesar di Indonesia yang menyediakan ulasan berdasarkan pengalaman langsung pengguna sehingga banyak dimanfaatkan sebagai sumber data dalam penelitian analisis sentimen produk skincare (Amalia et al., 2023; Wijayatun Pratiwi et al., 2021; Yutika et al., 2021).

Proses *scraping* dilakukan menggunakan bahasa pemrograman Python untuk memperoleh data ulasan pengguna secara otomatis dan terstruktur. Hasil proses *scraping* berhasil mengumpulkan sebanyak 1.029 ulasan pengguna yang selanjutnya digunakan sebagai dataset penelitian. Dataset tersebut kemudian melalui proses pelabelan sentimen, preprocessing, dan klasifikasi.

3.2. Hasil Pelabelan Sentimen (Ground Truth)

Setelah proses pengumpulan data selesai dilakukan, seluruh ulasan diberikan label sentimen secara manual (*ground truth*) oleh dua anotator independen. Pelabelan dilakukan dengan mengelompokkan setiap ulasan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses pelabelan ini bertujuan untuk menghasilkan label sentimen yang lebih representatif terhadap isi ulasan pengguna sehingga dapat digunakan sebagai dasar dalam proses klasifikasi sentimen.



Gambar 2. Distribusi Sentimen Hasil Ground Truth

Berdasarkan Gambar 2, Sentimen positif mendominasi dataset dengan jumlah 881 ulasan (85,62%), sedangkan sentimen netral berjumlah 92 ulasan (8,94%) dan sentimen negatif sebanyak 56 ulasan (5,44%). Hasil tersebut menunjukkan bahwa mayoritas pengguna memberikan tanggapan positif terhadap produk Skintific 5X Ceramide Barrier Repair Moisture Gel. Dominasi sentimen positif mengindikasikan bahwa produk dinilai mampu memenuhi ekspektasi pengguna, terutama dalam menjaga kelembapan kulit dan membantu memperbaiki *skin barrier*. Temuan ini sejalan dengan penelitian Harnelia dan Saputra (2024), Widhiyanta et al. (2025), dan Amalia et al. (2023) yang menunjukkan bahwa ulasan produk skincare dan moisturizer pada umumnya didominasi oleh sentimen positif.

Meskipun demikian, masih ditemukan sentimen negatif sebesar 5,44% yang menunjukkan adanya pengguna yang mengalami ketidakcocokan terhadap produk. Selain itu, distribusi sentimen yang tidak seimbang (*imbalanced dataset*) karena dominasi kelas positif perlu diperhatikan pada tahap klasifikasi dan evaluasi model (Rahmadzani et al., 2025).

3.3. Hasil Cohen's Kappa

Setelah proses pelabelan sentimen selesai dilakukan oleh dua anotator independen, dilakukan pengukuran tingkat kesepakatan antar anotator menggunakan metode Cohen's Kappa. Pengukuran ini bertujuan untuk memastikan bahwa label sentimen yang digunakan dalam penelitian memiliki tingkat konsistensi yang baik dan tidak terjadi secara kebetulan (Rahmadzani et al., 2025).

Tabel 1. Hasil Validasi Pelabelan Menggunakan Cohen's Kappa

Keterangan	Jumlah Data
Jumlah Data	1.029
Jumlah Perbedaan Label	106
Nilai Cohen's Kappa	0,6638
Interpretasi	Tinggi

Berdasarkan Tabel 1, diperoleh nilai Cohen's Kappa sebesar 0,6638 dengan jumlah perbedaan label sebanyak 106 ulasan. Nilai tersebut berada pada rentang 0,61–0,80 yang menunjukkan tingkat kesepakatan tinggi antar anotator. Hasil ini mengindikasikan bahwa proses pelabelan manual yang dilakukan memiliki reliabilitas yang baik sehingga layak digunakan sebagai dasar dalam proses klasifikasi sentimen. Meskipun masih ditemukan perbedaan label pada sebagian ulasan, seluruh perbedaan tersebut telah diselesaikan melalui proses diskusi untuk menentukan *gold label* yang digunakan pada penelitian ini. Dengan demikian, kualitas label sentimen yang digunakan dapat dianggap cukup representatif untuk menggambarkan opini pengguna terhadap produk yang diteliti.

3.4. Hasil Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan dan menormalisasi data teks sebelum dilakukan ekstraksi fitur menggunakan TF-IDF. Tahapan yang diterapkan meliputi *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*.

Tabel 2. Contoh Komentar yang telah di Preprocessing

Tahapan		Hasil Transformasi Teks	Keterangan / Diubah Menjadi
Komentar (Raw Text)	Asli	"salah satu moisturizer yang gue suka juga. nyaman dipake buat seluruh tipe kulit. teksturnya gel cream"	Data mentah hasil web scraping dari platform Female Daily.
1. Case Folding		"salah satu moisturizer yang gue suka juga nyaman dipake buat seluruh tipe kulit teksturnya gel cream"	Menghapus tanda baca, simbol, karakter khusus, serta mengubah seluruh huruf menjadi huruf kecil.
2. Normalisasi		"salah satu moisturizer yang saya suka juga nyaman pakai buat seluruh tipe kulit teksturnya gel cream"	Mengubah kata tidak baku menjadi kata baku, seperti "gue" menjadi "saya" dan "dipake" menjadi "pakai".
3. Tokenizing		['salah', 'satu', 'moisturizer', 'gue', 'suka', 'nyaman', 'dipake', 'seluruh', 'tipe', 'kulit']	Memecah kalimat menjadi kumpulan token atau kata tunggal.
4. Filtering (Stopword)		['moisturizer', 'suka', 'nyaman', 'dipake', 'tipe', 'kulit', 'tekstur', 'gel', 'cream']	Menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis sentimen.
5. Stemming		['moisturizer', 'suka', 'nyaman', 'pakai', 'tipe', 'kulit', 'tekstur', 'gel', 'cream']	Mengubah kata berimbuhan menjadi bentuk kata dasar agar variasi kata dapat diminimalkan.

Berdasarkan Tabel 2, terlihat bahwa setiap tahapan preprocessing mampu mengurangi variasi kata yang tidak diperlukan sehingga menghasilkan representasi teks yang lebih bersih dan terstruktur.

3.5. Hasil TF-IDF

Setelah melalui tahap *preprocessing*, data teks diubah ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Pembobotan TF-IDF bertujuan untuk mengidentifikasi kata-kata yang memiliki tingkat kepentingan tinggi dalam dataset sehingga dapat digunakan sebagai fitur pada proses klasifikasi.

Tabel 3. Hasil Kata dengan Bobot TF-IDF Tertinggi

Kata	Bobot
Kulit	0,0786
Cocok	0,0411
Pakai	0,0411
Produk	0,0401

Bagus

0,0392

Berdasarkan Tabel 3, kata *kulit* memiliki bobot TF-IDF tertinggi dibandingkan kata lainnya. Hal ini menunjukkan bahwa sebagian besar ulasan pengguna berfokus pada kondisi kulit setelah penggunaan produk. Selain itu, kata *cocok*, *pakai*, *produk*, dan *bagus* juga memiliki bobot yang tinggi sehingga menjadi fitur penting dalam proses klasifikasi sentimen.

3.6. Hasil Klasifikasi Naïve Bayes

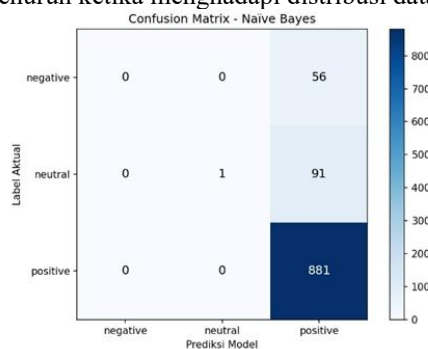
Kinerja algoritma Naïve Bayes dievaluasi menggunakan metode Stratified 10-Fold Cross Validation dengan metrik *accuracy*, *precision*, *recall*, dan *macro F1-score*.

Tabel 4. Hasil Klasifikasi Naïve Bayes

Metrik	Nilai
Accuracy	0,8562
Precision	0,6190
Recall	0,3370
Macro F1-score	0,3148

Berdasarkan Tabel 4, algoritma Naïve Bayes memperoleh nilai *accuracy* sebesar 85,62%. Meskipun nilai akurasi tergolong tinggi, nilai *recall* dan *macro F1-score* yang diperoleh masih relatif rendah, yaitu masing-masing sebesar 0,3370 dan 0,3148. Hasil tersebut menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas, khususnya sentimen netral dan negatif.

Berdasarkan *classification report*, model cenderung mengklasifikasikan sebagian besar ulasan ke dalam kategori positif. Kondisi ini diduga dipengaruhi oleh distribusi dataset yang tidak seimbang (*imbalanced dataset*), di mana sentimen positif mendominasi sebanyak 85,62% dari keseluruhan data. Akibatnya, model memiliki kecenderungan untuk memprediksi kelas mayoritas dibandingkan kelas lainnya. Hasil penelitian ini sejalan dengan penelitian Amalia et al. (2023) yang menunjukkan bahwa Naïve Bayes mampu menghasilkan akurasi yang baik pada klasifikasi sentimen, namun performanya dapat menurun ketika menghadapi distribusi data yang tidak seimbang.



Gambar 3. Confusion Matrix Naïve Bayes

Berdasarkan Gambar 3, Model Naïve Bayes berhasil mengklasifikasikan 881 ulasan positif dengan benar. Namun, model tidak mampu mengidentifikasi ulasan negatif dan hanya mampu mengklasifikasikan 1 ulasan netral dengan benar. Sebagian besar ulasan netral dan seluruh ulasan negatif diprediksi sebagai sentimen positif. Hasil ini menunjukkan bahwa model memiliki kecenderungan yang sangat kuat terhadap kelas mayoritas sehingga kemampuan dalam mengenali kelas minoritas masih rendah.

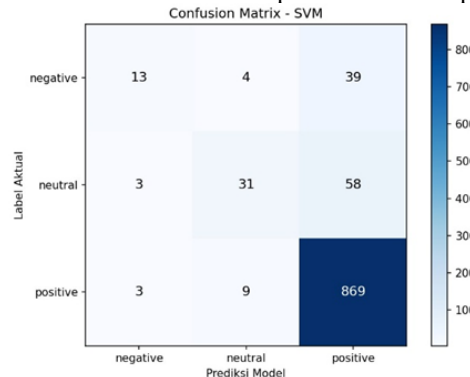
3.7. Hasil Klasifikasi SVM

Selain menggunakan Naïve Bayes, penelitian ini juga menerapkan algoritma SVM dengan kernel linear untuk melakukan klasifikasi sentimen.

Tabel 5. Hasil Klasifikasi SVM

Metrik	Nilai
Accuracy	0,8873
Precision	0,7628
Recall	0,5185
Macro F1-score	0,5812

Berdasarkan Tabel 5, algoritma SVM menghasilkan nilai *accuracy* sebesar 88,73%, lebih tinggi dibandingkan Naïve Bayes. Selain itu, nilai *precision*, *recall*, dan *macro F1-score* yang diperoleh juga menunjukkan peningkatan yang cukup signifikan. Nilai *macro F1-score* sebesar 0,5812 menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam mengenali seluruh kategori sentimen, termasuk kelas netral dan negatif yang jumlah datanya relatif sedikit. Hal ini menunjukkan bahwa SVM mampu membentuk batas pemisah (*hyperplane*) yang lebih optimal pada data teks hasil pembobotan TF-IDF. Hasil penelitian ini sejalan dengan penelitian Pratiwi et al. (Wijayatun Pratiwi et al., 2021) Harnelia dan Saputra (2024), serta Widhiyanta et al. (2025) yang menunjukkan bahwa SVM merupakan salah satu algoritma yang efektif untuk klasifikasi sentimen pada data ulasan produk skincare.



Gambar 4. Confusion Matrix SVM

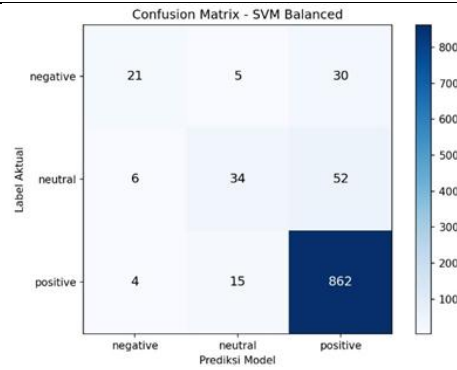
Model SVM berhasil mengklasifikasikan 869 ulasan positif, 31 ulasan netral, dan 13 ulasan negatif dengan benar. Dibandingkan Naïve Bayes, model SVM menunjukkan kemampuan yang lebih baik dalam mengenali kelas netral dan negatif sehingga menghasilkan nilai *recall* dan *macro F1-score* yang lebih tinggi.

3.8. Hasil Klasifikasi SVM Balanced

Tabel 6. Hasil Klasifikasi SVM Balanced

Metrik	Nilai
Accuracy	0,8912
Precision	0,7401
Recall	0,5743
Macro F1-score	0,6311

Hasil pengujian menunjukkan bahwa SVM Balanced menghasilkan *accuracy* sebesar 89,12%, *precision* sebesar 74,01%, *recall* sebesar 57,43%, dan *macro F1-score* sebesar 63,11%. Nilai tersebut lebih tinggi dibandingkan SVM standar yang memperoleh *accuracy* sebesar 88,73% dan *macro F1-score* sebesar 58,12%.



Gambar 5. Confusion Matrix SVM Balanced

Berdasarkan *confusion matrix*, SVM Balanced mampu mengklasifikasikan 862 ulasan positif, 34 ulasan netral, dan 21 ulasan negatif dengan benar. Peningkatan jumlah prediksi yang benar pada kelas negatif dan netral menunjukkan bahwa penerapan *class balancing* membantu model dalam mengenali kelas minoritas secara lebih baik. Hal ini berdampak pada peningkatan nilai *recall* dan *macro F1-score*, sehingga menghasilkan performa klasifikasi yang lebih seimbang dibandingkan SVM standar.

3.9. Visualisasi Word Cloud Berdasarkan Kategori Sentimen

a) Word Cloud Ulasan Positif



Gambar 6. Word Cloud Ulasan Positif

Kata-kata yang paling dominan pada sentimen positif antara lain “bagus”, “kulit”, “moisturizer”, “skin barrier,” “cocok”, “lembab”, dan “skintific”. Dominasi kata-kata tersebut menunjukkan bahwa sebagian besar pengguna merasa puas terhadap produk yang digunakan. Ulasan positif umumnya berkaitan dengan kemampuan produk dalam menjaga kelembapan kulit, memperbaiki “skin barrier”, serta memberikan kenyamanan saat digunakan. Hasil ini mendukung temuan pada distribusi sentimen yang menunjukkan bahwa mayoritas ulasan termasuk ke dalam kategori positif.

b) Word Cloud Ulasan Negatif



Gambar 7. Word Cloud Ulasan Negatif

Kata-kata yang sering muncul pada sentimen negatif antara lain “kulit”, “cocok”, “breakout”, “bruntusan”, “jerawat”, “kurang”, dan “muka”. Kemunculan kata-kata tersebut menunjukkan bahwa sebagian pengguna mengalami ketidakcocokan setelah menggunakan produk. Keluhan yang ditemukan umumnya berkaitan dengan munculnya jerawat, bruntusan, atau hasil penggunaan yang tidak sesuai dengan harapan. Meskipun jumlah ulasan negatif relatif sedikit dibandingkan sentimen positif, hasil ini menunjukkan bahwa efektivitas produk dapat berbeda pada setiap jenis kulit.

c) Word Cloud Ulasan Netral

DOI: <https://doi.org/10.69693/ijmst.v4i2.11777>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 8. Word Cloud Ulasan Netral

Kata-kata yang dominan pada sentimen netral antara lain “lembab”, “moisturizer”, “kulit”, “hidrasi,” “tekstur”, “ceramide”, dan “cocok”. Berbeda dengan sentimen positif maupun negatif, ulasan netral cenderung berisi informasi deskriptif mengenai kandungan, tekstur, manfaat, dan karakteristik produk tanpa menunjukkan kecenderungan sentimen yang kuat. Hal ini menunjukkan bahwa pengguna lebih banyak memberikan informasi atau pengalaman penggunaan secara objektif dibandingkan memberikan penilaian yang bersifat positif maupun negatif.

3.10. Perbandingan Performa Model

Tabel 7. Perbandingan Performa Model

Model	Accuracy	Precision	Recall	Macro F1-score
Naïve Bayes	85,62%	61,90%	33,70%	31,48%
SVM	88,73%	76,28%	51,85%	58,12%
SVM Balanced	89,12%	74,01%	57,43%	63,11%

Model SVM Balanced menghasilkan performa terbaik dengan nilai *accuracy* sebesar 89,12% dan *macro F1-score* sebesar 63,11%. Sementara itu, model SVM memperoleh *accuracy* sebesar 88,73% dan *macro F1-score* sebesar 58,12%, sedangkan Naïve Bayes menghasilkan *accuracy* sebesar 85,62% dan *macro F1-score* sebesar 31,48%. Hasil tersebut menunjukkan bahwa algoritma SVM mampu mengklasifikasikan sentimen ulasan dengan lebih baik dibandingkan Naïve Bayes. Selain itu, penerapan *class balancing* pada SVM berhasil meningkatkan kemampuan model dalam mengenali kelas minoritas, yang ditunjukkan oleh peningkatan nilai *recall* dan *macro F1-score*. Temuan ini sejalan dengan penelitian Pratiwi et al. (Wijayatun Pratiwi et al., 2021), Rahmadzani et al. (Rahmadzani et al., 2025), dan Widhiyanta et al. (2025) yang menunjukkan bahwa SVM memiliki performa yang lebih baik dibandingkan Naïve Bayes pada klasifikasi sentimen berbasis teks.

4. Kesimpulan

Penelitian ini berhasil melakukan analisis sentimen terhadap 1.029 ulasan produk Skintific 5X Ceramide Barrier Repair Moisture Gel pada platform Female Daily menggunakan algoritma Naïve Bayes dan SVM. Hasil pelabelan manual (*ground truth*) menunjukkan bahwa sentimen positif mendominasi dataset dengan jumlah 881 ulasan (85,62%), diikuti sentimen netral sebanyak 92 ulasan (8,94%) dan sentimen negatif sebanyak 56 ulasan (5,44%). Validasi pelabelan menggunakan Cohen’s Kappa menghasilkan nilai 0,6638 yang menunjukkan tingkat kesepakatan tinggi antar anotator sehingga label yang digunakan dinilai cukup reliabel untuk proses klasifikasi.

Berdasarkan hasil pengujian, algoritma Naïve Bayes memperoleh nilai *accuracy* sebesar 85,62% dengan *macro F1-score* sebesar 31,48%. Sementara itu, algoritma SVM menghasilkan performa yang lebih baik dengan *accuracy* sebesar 88,73% dan *macro F1-score* sebesar 58,12%. Penerapan *class balancing* pada SVM memberikan hasil terbaik dengan *accuracy* sebesar 89,12% dan *macro F1-score* sebesar 63,11%. Hasil tersebut menunjukkan bahwa SVM lebih efektif dibandingkan Naïve Bayes dalam mengklasifikasikan sentimen ulasan produk skincare, terutama pada dataset yang memiliki distribusi kelas tidak seimbang. Dengan demikian, SVM Balanced dapat direkomendasikan sebagai model terbaik dalam penelitian ini karena mampu menghasilkan performa klasifikasi yang lebih seimbang pada seluruh kategori sentimen.

Temuan penelitian ini menunjukkan bahwa penggunaan *ground truth* manual yang divalidasi menggunakan Cohen’s Kappa mampu menghasilkan dataset yang lebih reliabel untuk proses klasifikasi sentiment. Penelitian selanjutnya dapat mempertimbangkan penggunaan deep learning seperti IndoBERT atau RoBERTa serta teknik penanganan data tidak seimbang lainnya agar performa klasifikasi semakin meningkat.

DOI: <https://doi.org/10.69693/ijmst.v4i2.11777>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Referensi

- Amalia, D., Totohendarto, M. H., & Alam, S. (2023). Analisis Sentimen Produk Populer Moisturizer Pada Female Daily Menggunakan Metode Naive Bayes. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, 8(2), 108. <https://doi.org/10.51211/itbi.v8i2.2483>
- Clarisha, W., Fani, A. A. M., Surianto, D. F., & Fadilah, N. (2025). Analisis Sentimen Sunscreen Lokal Skintific, Somethinc, dan Avoskin dengan Naive Bayes dan SVM. *Jambura Journal of Electrical and Electronics Engineering*, 7(2), 264–271.
- Harmandini, K. P., & L, K. M. (2024). Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment. *Sinkron*, 8(2), 929–937. <https://doi.org/10.33395/sinkron.v8i2.13376>
- Harnelia, H. (2024). Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (Svm). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2). <https://doi.org/10.23960/jitet.v12i2.4095>
- Jannah, N. Z. B., & Kusnawi, K. (2024). Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces. *Sinkron*, 8(2), 727–733. <https://doi.org/10.33395/sinkron.v8i2.13559>
- Khoerunnisa, S., Shiddiq, D. F., & Nurhayati, D. (2025). Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 566–577. <https://doi.org/10.57152/malcom.v5i2.1852>
- Khotimah, A. K., & Arianto, D. B. (2025). Analisis Sentimen Ulasan Produk Populer Sunscreen di Female Daily Menggunakan Metode Naive Bayes. 5(2), 54–60.
- Lestari, V. B., & Hutagalung, C. A. (2025). Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes. *Journal of Computer Science and Applications*, 8(1), 22–2025.
- Melisa Nur Aini, Rita Yulfani, & Nurul Jariah. (2024). Application of the Naive Bayes Method for Sentiment Analysis of Sunscreen Product Reviews Based on the Female Daily Review. *Jurnal Multimedia Dan Teknologi Informasi (Jatilima)*, 6(01), 24–34.
- Rahmadzani, R., Pratiwi, R., & Paradhita, A. (2025). Comparison of Naive Bayes and Support Vector Machine (SVM) Methods in Female Daily Skincare Sentiment Analysis Perbandingan Metode Naive Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Skincare Female Daily. *Journal of Applied Social Dan Education Studies*, 6(2), 99–108.
- Widhiyanta, N., Muhandhis, I., Jannah, R. S., & Wulansari, S., Dairoh, Intan Af'idah, D., Rifa A, Q., & Gian F, A. (2021). Analisis Sentimen Pada Review Skincare Female Daily menggunakan SVM. *Journal of Informatics, Information System, Software Engineering and Applications*, 1, 40–46.
- Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422. <https://doi.org/10.30865/mib.v5i2.2845>