

## **Comparison of Naive Bayes and Support Vector Machine (SVM) Methods in Female Daily Skincare Sentiment Analysis**

RIFQI FAUZI RAHMADZANI<sup>1</sup>, RISZKI WIJAYATUN PRATIWI<sup>2</sup>, ASTRID NOVIANA PARADHITA<sup>3</sup>

<sup>1</sup>*Universitas Batik Surakarta, Central Java, Indonesia*  
[rifqifai@gmail.com](mailto:rifqifai@gmail.com)

<sup>2</sup>*Politeknik Assalaam Surakarta, Central Java, Indonesia*  
[riszkiwp@gmail.com](mailto:riszkiwp@gmail.com)

<sup>3</sup>*Universitas Sebelas Maret, Central Java, Indonesia*  
[astrid.noviana@staff.uns.ac.id](mailto:astrid.noviana@staff.uns.ac.id)

### **Abstract**

The development of the beauty industry in Indonesia has increased significantly, along with the high participation of consumers in providing online reviews, especially through the Female Daily platform. These reviews contain opinions that can be analyzed to determine consumer sentiment towards a product. This study seeks to perform sentiment analysis on beauty product reviews by applying two text classification techniques: Support Vector Machine (SVM) and Naive Bayes. The research stages include collecting review data from Female Daily, text preprocessing (such as tokenization, stemming, and stopword removal), and sentiment classification into two categories, namely Yes and No. The evaluation results indicate that the SVM method outperforms Naive Bayes, achieving a higher level of accuracy. SVM is able to capture more complex patterns in text data, while Naive Bayes tends to produce lower performance due to the assumption of independence between features. The evaluation results demonstrate that the SVM method performs better than Naive Bayes, achieving higher accuracy scores. SVM excels at recognizing more intricate patterns in textual data, whereas Naive Bayes often shows lower performance due to its assumption of feature independence. Overall, the majority of user reviews are positive, reflecting satisfaction with certain beauty products. This study shows that the SVM method is more recommended for text-based sentiment analysis of reviews in the beauty industry, especially in the context of diverse and unstructured consumer review data. As for the accuracy results of the Naive Bayes method is 80% and the accuracy results of the SVM method is 87%.

**Keywords:** sentiment analysis, female daily, beauty products, support vector machine, naive bayes, text classification



Copyright © 2025 The Author(s)

This is an open-access article under the CC BY-SA license.

## Perbandingan Metode Naive Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Skincare Female Daily

### Abstrak

Perkembangan industri kecantikan di Indonesia mengalami peningkatan signifikan, seiring dengan tingginya partisipasi konsumen dalam memberikan ulasan secara daring, khususnya melalui platform *Female Daily*. Ulasan-ulasan tersebut berisi pendapat yang dapat dianalisis untuk mengidentifikasi sentimen konsumen terhadap sebuah produk. Penelitian ini bertujuan untuk melakukan analisis sentimen pada ulasan produk kecantikan dengan menggunakan dua metode klasifikasi teks, yaitu *Support Vector Machine (SVM)* dan *Naive Bayes*. Tahapan penelitian meliputi pengumpulan data *review* dari *Female Daily*, *preprocessing* teks (seperti tokenisasi, *stemming*, dan *stopword removal*), serta klasifikasi sentimen ke dalam dua kategori, yaitu Ya dan Tidak. Hasil evaluasi mengindikasikan bahwa metode SVM memberikan kinerja yang lebih unggul dibandingkan Naive Bayes, ditunjukkan oleh nilai akurasi yang lebih tinggi. SVM mampu mengenali pola-pola kompleks dalam data teks, sedangkan Naive Bayes cenderung memiliki kinerja lebih rendah karena mengasumsikan adanya independensi antar fitur. Secara keseluruhan, mayoritas ulasan pengguna bersentimen positif, yang mencerminkan kepuasan terhadap produk kecantikan tertentu. Penelitian ini menunjukkan bahwa metode SVM lebih direkomendasikan untuk analisis sentimen berbasis teks ulasan di industri kecantikan, terutama dalam konteks data ulasan konsumen yang beragam dan tidak terstruktur. Adapun untuk hasil Akurasi dari metode *Naive Bayes* adalah 80% dan untuk hasil akurasi dari metode SVM adalah 87%.

**Kata kunci:** analisis sentimen, *female daily*, produk kecantikan, *support vector machine*, *naive bayes*, klasifikasi teks

### PENDAHULUAN

Perkembangan teknologi informasi saat ini telah mendorong masyarakat untuk berbagi pengalaman dan opini mereka melalui platform digital (Aljabar dan Karomah, 2024). Salah satu platform yang sangat populer di Indonesia, khususnya di bidang kecantikan, adalah *Female Daily*. Di platform ini, pengguna dapat memberikan ulasan mengenai berbagai produk skincare. Ulasan yang disampaikan oleh pengguna tersebut mengandung informasi berharga yang bisa dianalisis untuk memahami persepsi konsumen terhadap produk-produk tertentu (Putri & Ardiansyah, 2023).

Analisis sentimen adalah proses dalam lingkup pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan untuk mendeteksi serta

mengelompokkan opini dalam bentuk teks ke dalam kategori positif atau negatif (Kaharudin et al., 2023). Teknologi NLP telah banyak dimanfaatkan sebagai alat untuk memahami pandangan publik terkait produk atau layanan tertentu (Afidah et al., 2022).

Analisis sentimen dalam industri skincare memiliki urgensi dan relevansi yang semakin meningkat seiring dengan pertumbuhan signifikan pasar kecantikan, baik secara global maupun di Indonesia. Konsumen kini semakin sadar akan pentingnya perawatan kulit, sehingga mereka lebih selektif dalam memilih produk yang sesuai dengan kebutuhan kulit mereka. Hal ini mendorong merek-merek kecantikan untuk memahami persepsi dan preferensi konsumen melalui ulasan yang ditinggalkan di platform komunitas seperti *Female Daily*. Dengan menganalisis sentimen dari ulasan-ulasan tersebut, perusahaan dapat memperoleh wawasan strategis mengenai kelebihan, kekurangan, dan ekspektasi konsumen terhadap produk mereka, serta meningkatkan kualitas produk dan strategi pemasaran.

Namun, menganalisis sentimen dalam konteks ulasan skincare memiliki tantangan tersendiri. Bahasa yang digunakan dalam ulasan konsumen sering kali informal, mengandung istilah-istilah spesifik kecantikan, serta dipengaruhi oleh konteks sosial dan budaya lokal. Oleh karena itu, dibutuhkan metode analisis sentimen yang mampu menangani kompleksitas bahasa tersebut secara akurat. Dalam hal ini, metode klasifikasi seperti *Naive Bayes* dan *Support Vector Machine* (SVM) menjadi pilihan yang banyak digunakan dalam penelitian karena efektivitas dan efisiensinya dalam menangani data teks yang besar dan beragam.

Beberapa penelitian terbaru mendukung pentingnya pendekatan ini. Lestari et al. (2024) membandingkan efektivitas berbagai kernel SVM dan pendekatan berbasis leksikon dalam analisis sentimen ulasan produk skincare, dengan hasil yang menunjukkan bahwa pemilihan metode sangat berpengaruh terhadap akurasi klasifikasi. Penelitian oleh Nabila dan Putra (2024) juga menggarisbawahi pentingnya analisis sentimen terhadap merek skincare yang viral, mengidentifikasi berbagai faktor emosional yang mempengaruhi keputusan pembelian. Sementara itu, Rambe et al. (2023) mengeksplorasi penggunaan metode Naive Bayes dalam konteks aplikasi produk kecantikan, dan menemukan bahwa pendekatan ini cukup efektif dalam memetakan sentimen konsumen. Temuan-temuan ini menunjukkan bahwa perbandingan antara metode *Naive Bayes* dan SVM dalam konteks ulasan *skincare* bukan hanya relevan secara akademis, tetapi juga aplikatif bagi industri.

Beragam metode klasifikasi telah diterapkan untuk meningkatkan akurasi dalam analisis sentimen. Dua metode yang cukup dikenal adalah *Naive Bayes* dan

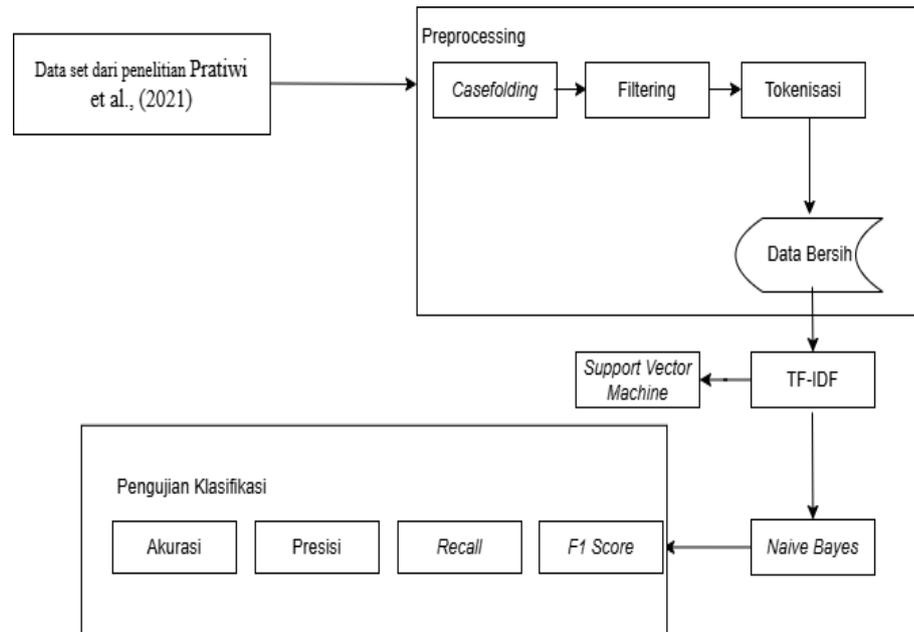
*Support Vector Machine* (SVM). *Naive Bayes* merupakan metode klasifikasi berbasis probabilistik yang menggunakan teorema Bayes, dengan asumsi bahwa setiap fitur bersifat independen satu sama lain (Rachman & Handayani, 2021), Sementara itu, SVM adalah metode klasifikasi yang berbasis pada margin. Metode ini berfungsi dengan mencari *hyperplane* terbaik yang dapat memisahkan data dari kelas-kelas yang berbeda. (Prasetyo et al., 2022).

Beberapa penelitian sebelumnya telah membandingkan performa kedua metode ini dalam analisis sentimen. Sebagai contoh, penelitian yang dilakukan oleh Pratiwi et al., (2021) menerapkan metode SVM untuk menganalisis sentimen pada ulasan skincare di *Female Daily*, dan berhasil mencapai akurasi sebesar 87%. Sementara itu, penelitian Mustikananda et al., (2022) membandingkan metode *Naive Bayes* dan SVM dalam analisis sentimen terhadap ulasan produk kosmetik, serta menemukan bahwa SVM menghasilkan akurasi yang lebih tinggi daripada *Naive Bayes*.

Namun, kinerja metode klasifikasi dapat bervariasi secara signifikan, tergantung pada karakteristik data dan konteks penggunaannya. Oleh karena itu, sangat penting untuk mengevaluasi kedua metode ini dalam kasus yang spesifik, seperti ulasan produk skincare yang biasanya ditulis dengan gaya bahasa informal dan sifat subyektif. Tujuan dari penelitian ini adalah untuk membandingkan performa metode *Naive Bayes* dan SVM dalam melakukan analisis sentimen pada ulasan produk *skincare* yang terdapat di platform *Female Daily*. Penelitian ini memanfaatkan dataset dari studi yang dilakukan oleh penelitian Pratiwi et al. (2021), dan hasil dari analisis sentimen akan dievaluasi melalui matriks penilaian yang mencakup akurasi, presisi, dan recall.

## **METODE PENELITIAN**

Data yang dipakai dalam studi ini berasal dari dataset yang dikumpulkan oleh Pratiwi et al. (2021), dengan total 1249 ulasan dari situs *female daily* yang diperoleh melalui teknik pengambilan data otomatis. Di dalamnya terdapat 649 ulasan positif dan 600 ulasan negatif. Selanjutnya, data ini dibagi menjadi data untuk pelatihan dan pengujian. Rincian proses analisis sentimen dapat dilihat dalam Gambar 1.



Gambar 1. Proses Sentimen Analisis dengan Metode SVM dan *Naïve Bayes*

### *Preprocessing*

Dalam penelitian ini, kami menggunakan data dari studi sebelumnya, yaitu dari Pratiwi et al. (2021), dimana dataset ini sudah dilabeli. Selanjutnya, tahapan yang dilakukan adalah preprocessing, yang bertujuan untuk membersihkan dataset agar siap untuk proses klasifikasi. Proses preprocessing dalam penelitian ini mencakup *casefolding*, *filtering*, dan tokenisasi. *Casefolding* bertujuan untuk mengubah semua huruf dalam ulasan menjadi huruf kecil. Setelah itu, *filtering* dilakukan untuk menghapus karakter spesifik seperti \$, #, %, , dan lain-lain, serta untuk menghapus kata-kata tidak relevan yang dihasilkan dari parsing, seperti username @, hashtag, URL, dan emotikon. Selain itu, simbol, tanda baca, dan angka juga dihilangkan karena tidak berpengaruh signifikan pada penentuan label. Langkah terakhir dalam preprocessing adalah tokenisasi, yang digunakan untuk memecah kalimat menjadi kata. Proses ini dilakukan dengan memisahkan kata-kata berdasarkan spasi dalam kalimat.

### *Term Frequency inverse-document frequency (TF-IDF)*

TF-IDF merupakan kombinasi dua algoritma, yaitu Frekuensi dan Frekuensi Dokumen Terbalik. Frekuensi Istilah menunjukkan seberapa banyak istilah muncul di dalam suatu dokumen, sementara frekuensi dokumen menghitung seberapa sering istilah muncul di berbagai dokumen, dengan cara membalikkan frekuensi dokumen yang termasuk suatu kata. Pada penelitian *tools* yang digunakan

menggunakan *google colab* dengan menggunakan bahasa pemrograman *python* dan beberapa librarynya yaitu *sklearn* serta *Gensim*.

#### Metode SVM & *Naïve Bayes*

Data yang telah melalui tahap *preprocessing* maka akan dibagi menjadi dua kategori, yaitu data untuk pelatihan dan data untuk pengujian, dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Setelah membagi data pelatihan dan pengujian, langkah selanjutnya adalah melakukan pembobotan kata dengan menggunakan TF-IDF, lalu dilanjutkan dengan proses klasifikasi yang menerapkan dua metode. Metode klasifikasi pertama yang digunakan adalah SVM, diikuti oleh klasifikasi kedua yang memakai metode *Naïve Bayes*. Dalam penelitian ini, tujuan penggunaan metode tersebut adalah untuk mengevaluasi kinerja terbaik dari machine learning klasik, yaitu SVM dan *Naïve Bayes*. Kinerja kedua metode ini akan diuji menggunakan metrik berupa *akurasi*, *presisi*, *recall*, dan *f-measure*.

### HASIL DAN PEMBAHASAN

Data yang digunakan berjumlah 1249 ulasan, pada penelitian data ulasan dibagi menjadi dua jenis yaitu positif dan negatif. Berdasarkan hasil eksperimen yang sudah dilakukan pada penelitian ini dengan menggunakan SVM dapat menghasilkan akurasi sebesar 87%, hal ini sama dengan penelitian (Pratiwi et al., 2021) menggunakan data ulasan dari *female daily* dan metode SVM menghasilkan akurasi sebesar 87%. Di sisi lain, penelitian ini juga menambahkan metode perbandingan, yaitu *naïve bayes*, dengan dataset yang sama dan membagi data menjadi 80% untuk data pelatihan dan 20% untuk data pengujian, yang menghasilkan akurasi sebesar 80. Penelitian ini memanfaatkan data dari studi sebelumnya (Pratiwi et al. , 2021) dan menerapkan metode SVM serta *naïve bayes*, sehingga dapat mengevaluasi kinerja terbaik dari kedua metode tersebut. Pembagian data pelatihan adalah 80% dan data pengujian 20%. Contoh dataset yang dipakai pada penelitian ini dapat dilihat di Tabel 1.

Tabel 1 Ulasan pada Website *Female Daily*

No	Ulasan	Kelas Sentimen
1	Teksturnya gel gitu jadi cepat terserap ke dalam kulit ga ninggalin efek basah atau lengket dikitpun komposisinya juga bagus banget jadi bisa dianggap sunscreen dan skincare sekaligus gitu, intinya bagus banget dan membuat kulitku mulus dan terhidrasi	Ya
2	Suka banget teksturenya serumnya ringan cepat meresap ke kulit jadi gak ngeganggu pake serum di siang hari, aromanya juga seger sih menurut aku, jadi tinggal tetesin aja ke jerawat besokannya udah kering, rekomended banget deh buat kulit yang gampang jerawat	Ya
3	Agak kurang suka sih sama teksturnya dan after pemakaiannya. Bikin muka ngerasa dempul gitu whiteningta juga ga ngefek di muka aku walaupun sebenarnya seneng karena spfnya 35 yang ngebuat cream ini ngelindungi paparan sinar matahari	Tidak
4	Pernah nyoba produk ini karena adanya ini di took deket rumah, cuma sekali pakai maaf banget langsung muncul jerawat sepertinya memang tidak cocok. Kelebihannya sih gampang dibeli di took mana aja dan ukurannya ekonomis	Tidak

Data yang telah melewati proses *preprocessing* selanjutnya akan menjalani tahap klasifikasi analisis sentimen dengan membandingkan dua metode. Yaitu, SVM dan *naïve bayes*. Metode pengujian yang digunakan adalah confusion matrix. Ini merupakan suatu cara yang digunakan dalam menghitung nilai akurasi pada data mining. Rincian tabelnya bisa dilihat di Tabel 2. Selanjutnya, hasil perbandingan antara kedua metode, baik SVM maupun *naïve bayes*, dalam penelitian ini dapat ditemukan di Tabel III.

Tabel II *Confusion Matrix*

<i>Predicted Values</i>	<i>Actual Values</i>	
	<i>1 (Positive)</i>	<i>0 (Negative)</i>
<i>1 (Positive)</i>	<i>TP (True Positive)</i>	<i>FP (False Positive)</i>
<i>0 (Negative)</i>	<i>FN (False Negative)</i>	<i>TN (True Negative)</i>

Tabel III Hasil Pengujian Perbandingan SVM dan Naïve Bayes

Parameter	Hasil Pengujian SVM	Hasil Pengujian Naïve Bayes
Data Training	80%	80%
Data Testing	20%	20%
Hasil Akurasi	87%	80%
Hasil Recall	90%	86%
Hasil Presisi	84%	75,65%
Hasil F1 Score	87%	79.58%

## SIMPULAN

Penelitian ini memiliki tujuan untuk melakukan analisis sentimen ulasan produk kecantikan yang diambil dari platform *Female Daily* dengan menggunakan dua metode klasifikasi, yaitu SVM dan *Naive Bayes*. Data yang dipakai merupakan kumpulan ulasan pengguna mengenai berbagai produk kecantikan, yang kemudian dikategorikan ke dalam dua kelas sentimen, yaitu Ya dan Tidak.

Berdasarkan hasil analisis dan pengujian yang telah dilakukan, dapat disimpulkan beberapa hal. Metode SVM menunjukkan kinerja yang lebih unggul dibandingkan *Naive Bayes* dalam mengklasifikasikan sentimen ulasan. Keunggulan ini tercermin dari nilai akurasi, presisi, recall, dan F1 Score yang lebih tinggi pada model SVM. SVM mampu menangkap pola-pola kompleks dalam data teks, sehingga lebih akurat dalam membedakan antara sentimen Ya dan Tidak. Sedangkan Metode *Naive Bayes*, walaupun lebih cepat dan ringan secara komputasi, memiliki keterbatasan dalam menangani data yang memiliki hubungan antar fitur yang kompleks. Hal ini memiliki dampak pada hasil klasifikasi yang cenderung kurang akurat dibandingkan SVM, terutama jika data ulasan mengandung ironi, sarkasme, atau konteks yang ambigu.

**DAFTAR PUSTAKA**

- Afidah, D. I., Dairoh, Handayani, S. F., & Pratiwi, R. W. (2022). Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, Vol. 21, No. 3, Juli 2022, p. 607~618.
- Aljabar, A., & Karomah, B. M. (2024). *Mengungkap Opini Publik: Pendekatan BERT-based- caused untuk Analisis Sentimen pada Komentar Film*. 5(1).
- Kaharudin, A., Supriyadi, A. A., Baitika, H., & Derryanur, M. (2023). *Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes: Kajian Literatur Review*. 2(6).
- Lestari, L. S., Sutrisno, T., & Lewenusa, I. (2024). Sentiment analysis on skincare product reviews using lexicon-based and comparison of SVM kernel. *Jurnal Indonesia Sosial Teknologi*, 5(11), 5250–5259.  
<https://jist.publikasiindonesia.id/index.php/jist/article/view/8769>
- Mustikananda, D., Ratnawati, D. E., & Rahayudi, B. (2023). *Perbandingan Algoritma Naïve Bayes dan Support Vector Machine untuk Analisis Sentimen terhadap Review Produk Aster Kosmetik Malang Marketplace Shopee*.
- Nabila, N. I. S., & Putra, E. D. (2024). Consumer sentiment analysis of viral skincare brands: A literature review. *Proceeding of International Seminar Conference of Economics and Business Excellence*.  
<https://conference.ut.ac.id/index.php/iscebe/article/view/4353>
- Prasetyo, T. M., Amrullah, A., Syahrir, S., & Sari, B. N. (2022). *Implementasi Algoritma Svm (Support Vector Machine) dalam Klasifikasi Penyakit Paru-Paru Berdasarkan Fitur Pola Bentuk*.
- Pratiwi, R. W., H, S. F., Dairoh, D., Afidah, D. I., A, Q. R., & F, A. G. (2021). Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM). *Journal of Informatics, Information System, Software*

*Engineering and Applications (INISTA)*, 4(1), 40–46.

<https://doi.org/10.20895/inista.v4i1.387>

Putri, N. A. R. & Ardiansyah. (2023). Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa. *Jurnal Sains dan Informatika*, 9(2), 136–145. <https://doi.org/10.34128/jsi.v9i2.649>

Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *Jurnal Informatika*, 8(2), 111–122. <https://doi.org/10.31294/ji.v8i2.10494>

Rambe, T. S., Hasibuan, M. N. S., & Dar, M. H. (2023). Sentiment analysis of beauty product applications using the Naïve Bayes method. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 7(2), 980–989.

<https://www.jurnal.polgan.ac.id/index.php/sinkron/article/view/12303>