

# Klasifikasi Sentimen Ulasan Sunscreen-1.docx

*by* Turnitin Student

---

**Submission date:** 09-Jul-2025 08:54AM (UTC+0700)

**Submission ID:** 2711911273

**File name:** Klasifikasi\_Sentimen\_Ulasan\_Sunscreen-1.docx (367.8K)

**Word count:** 2109

**Character count:** 14459

# KLASIFIKASI SENTIMEN ULASAN PRODUK SUNSCREEN PADA FEMALE DAILY MENGGUNAKAN METODE NAÏVE BAYES

Leonyka Grace Baktria, Baenil Huda, Elfina Novalia, Tukino

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat

[Si23.leonykabaktria@mhs.ubpkarawang.ac.id](mailto:Si23.leonykabaktria@mhs.ubpkarawang.ac.id), [baenilhuda@ubpkarawang.ac.id](mailto:baenilhuda@ubpkarawang.ac.id), [elfinanovalia@ubpkarawang.ac.id](mailto:elfinanovalia@ubpkarawang.ac.id), [tukino@ubpkarawang.ac.id](mailto:tukino@ubpkarawang.ac.id).

## Abstract

The development of internet technology encourages consumers to be more active in sharing their experiences through reviews, one of which is on the Female Daily platform. Sunscreen product reviews from users provide valuable sentiment insights. However, analyzing large-scale data manually is ineffective. This study aims to evaluate the sentiment of sunscreen product reviews using the Naïve Bayes Classifier algorithm. Data was collected through web scraping, followed by text pre-processing and sentiment labeling according to the ranking score into three categories: very suitable, suitable, and unsuitable. The imbalance in class distribution was addressed using the oversampling technique, and the data was then converted into a numeric format with TF-IDF. The model was trained with the Multinomial Naïve Bayes algorithm and evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The evaluation results showed that the model achieved an accuracy of 83.33%, with a precision of 0.84, a recall of 0.83, and an F1-score of 0.83. WordCloud visualization was used to identify dominant words in each sentiment category. These findings demonstrate the effectiveness of the Naïve Bayes algorithm in classifying consumer opinions well and highlight its potential for developing review-based product recommendation systems, as well as for understanding consumer perceptions in the beauty industry.

**Keywords:** Sentiment analysis, Naïve Bayes, Sunscreen, Female Daily, TF-IDF, Text classification

## Abstrak

Perkembangan teknologi internet mendorong konsumen untuk lebih aktif membagikan pengalamannya melalui ulasan, salah satunya pada platform Female Daily. Ulasan produk sunscreen dari pengguna memberikan wawasan sentimen yang berharga. Namun, menganalisis data dalam skala besar secara manual tidaklah efektif. Studi ini bertujuan untuk mengevaluasi sentimen ulasan produk tabir surya menggunakan algoritma *Naïve Bayes Classifier*. Data dikumpulkan melalui web scraping, diikuti oleh pra-pemrosesan teks dan pelabelan sentimen menurut skor peringkat menjadi tiga kategori: sangat cocok, cocok, dan tidak cocok. Ketidakseimbangan dalam distribusi kelas diatasi menggunakan teknik oversampling, dan data kemudian diubah menjadi format numerik dengan TF-IDF. Model dilatih dengan algoritma *Multinomial Naïve Bayes* dan dievaluasi menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model mencapai akurasi 83,33%, dengan presisi 0,84, recall 0,83, dan skor F1-score 0,83. Visualisasi WordCloud digunakan untuk mengidentifikasi kata-kata dominan di setiap kategori sentimen. Temuan ini menunjukkan efektivitas algoritma Naïve Bayes dalam mengklasifikasikan opini konsumen dengan baik dan menyoroti potensinya untuk mengembangkan sistem rekomendasi produk berbasis ulasan, serta untuk memahami persepsi konsumen dalam industri kecantikan.

**Kata Kunci:** Analisis sentimen, Naïve Bayes, Sunscreen, Female Daily, TF-IDF, Klasifikasi teks

## PENDAHULUAN

Munculnya platform media sosial dan kemajuan teknologi jaringan secara

fundamental telah memengaruhi perubahan sikap konsumen, termasuk dalam cara memperoleh informasi sebelum melakukan

pembelian. Platform ulasan daring seperti Female Daily telah menjadi sumber utama bagi konsumen untuk berbagi pengalaman sekaligus mendapatkan rekomendasi produk kecantikan. Di sektor kecantikan, ulasan produk perawatan kulit seperti *sunscreen* menjadi semakin relevan, seiring tumbuhnya kesadaran mengenai pentingnya melindungi kulit dari radiasi ultraviolet (UV). *Sunscreen* merupakan produk topikal yang berfungsi untuk melindungi kulit dari efek buruk sinar ultraviolet (UV), terutama sinar UVA dan UVB yang dapat menyebabkan penuaan dini, hiperpigmentasi, bahkan kanker kulit. Indonesia sebagai negara beriklim tropis dengan intensitas sinar UV yang tinggi sepanjang tahun, penggunaan *sunscreen* menjadi kebutuhan penting yang semakin disadari oleh masyarakat. Ulasan-ulasan pada Female Daily menghasilkan informasi sentimen yang berharga. Akan tetapi, volume data yang besar membuat analisis manual menjadi tidak efisien. Dalam konteks ini, pendekatan komputasional seperti analisis sentimen menjadi relevan. Analisis sentimen adalah studi sistematis yang bertujuan untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini atau emosi dalam sebuah teks. Salah satu metode yang sering digunakan dalam klasifikasi teks adalah *Naïve Bayes*, sebuah algoritma berbasis probabilitas yang sederhana namun efektif dalam mengelompokkan teks ke dalam kategori tertentu berdasarkan pola data yang ada. Penelitian serupa dilakukan oleh Zahirma dan Rumini (2025) yang menganalisis sentimen review pengguna yang memanfaatkan produk *sunscreen* Wardah UV Shield Essential Gel SPF 35 PA+++.

Dalam penelitian ini, data sebanyak 1.451 review dikumpulkan dari platform Female Daily, diolah melalui tahap text preprocessing dan transformasi TF-IDF, kemudian diklasifikasikan menggunakan algoritma *Naïve Bayes*. Hasil evaluasi model menunjukkan akurasi sebesar 79%, presisi sebesar 67%, recall sebesar 64%, dan F1-score sebesar 65%. Hasil ini mengindikasikan bahwa pendekatan *Naïve Bayes* berhasil dalam mengklasifikasikan sentimen, walaupun ada tantangan seperti ketidakseimbangan label dan anggapan independensi antara kata-kata yang

bisa berdampak pada ketepatan model. Berdasarkan urgensi tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen pada review produk *sunscreen* yang diperoleh dari platform Female Daily, dengan menerapkan algoritma *Naïve Bayes Classifier*. Data review dijelaskan melalui serangkaian tahapan komputasi, dimulai dari studi literatur dan pengumpulan data, kemudian dilanjutkan dengan text preprocessing dan sentiment labeling berdasarkan rating pengguna. Dalam penelitian ini, sentimen pengguna diklasifikasikan ke dalam tiga kelas utama, yaitu *sangat cocok*, *cocok*, dan *tidak cocok*. Klasifikasi multi-kelas ini dipilih untuk menangkap nuansa opini pengguna secara lebih detail dan realistis dibandingkan pendekatan biner. Pada fase berikutnya, informasi dipecah menjadi set data latihan dan set data pengujian. Mengingat adanya ketidakseimbangan distribusi antara kelas, terutama pada dominasi ulasan "cocok", dilakukan proses penambahan jumlah sampel pada data latihan dengan metode Random Over Sampler agar representasi setiap kelas menjadi seimbang. Langkah ini penting untuk mengurangi potensi bias model terhadap kelas mayoritas dan meningkatkan sensitivitas klasifikasi terhadap kelas minoritas seperti "tidak cocok" dan "sangat cocok". Data yang sudah diseimbangkan kemudian diubah menjadi format angka dengan menggunakan pendekatan TF-IDF, dan setelah itu dilatih memakai algoritma *Naïve Bayes*. Model dievaluasi melalui matriks kebingungan dengan memanfaatkan metrik akurasi, presisi, recall, dan skor F1. Selain itu, visualisasi WordCloud diterapkan untuk menemukan kata-kata yang paling sering muncul dalam setiap kategori sentimen. Penelitian ini diharapkan mampu berkontribusi pada pengembangan sistem klasifikasi opini berbasis teks dalam bahasa Indonesia, sekaligus menjadi referensi bagi pelaku industri kecantikan dalam memahami persepsi konsumen secara lebih komprehensif dan berbasis data.

## METODOLOGI PENELITIAN

### A. Skema Alur Penelitian



### F. Oversampling

Pada tahap ini, dilakukan proses penyeimbangan kelas (*class balancing*) terhadap label sentimen yang telah dihasilkan sebelumnya. Untuk mengatasi ketidakseimbangan ini, digunakan metode *oversampling* dengan pendekatan *random resampling*. Teknik ini bekerja dengan cara menggandakan (sampling dengan pengembalian) data dari kelas yang jumlahnya lebih sedikit hingga mencapai jumlah yang seimbang dengan kelas lainnya.

### G. Split Data

Setelah menyelesaikan tahap pelabelan dan penyeimbangan data, langkah berikutnya adalah membagi dataset menjadi dua kelompok utama: data latih yang mengambil 75% dan 25% sebagai data uji. Pemisahan ini bertujuan untuk membedakan data yang digunakan dalam proses pelatihan model dan data yang digunakan untuk mengevaluasi performa model.

### H. Pembobotan TF-IDF

Setiap tinjauan yang telah dibersihkan diubah menjadi vektor angka dengan memanfaatkan metode Term Frequency-Inverse Document Frequency (TF-IDF). Cara ini menggambarkan seberapa penting sebuah kata dalam dokumen dibandingkan dengan keseluruhan kumpulan teks, sehingga meningkatkan kemampuan model untuk mengenali pola sentimen.

### I. Pelatihan Model

Algoritma Multinomial Naive Bayes, yang sering digunakan untuk kategorisasi data teks, yang digunakan selama tahap klasifikasi. Model dilatih menggunakan data dari ekstraksi fitur TF-IDF dan label anotasi. Nilai parameter alpha ditetapkan sebesar 0,1 sebagai teknik pemulusan Laplace untuk menghindari probabilitas nol bagi kata-kata yang tidak muncul dalam data latih. Pemilihan nilai alpha dilakukan secara empiris dengan mencoba beberapa nilai kecil seperti 0,1 dan 0,3 untuk mengoptimalkan kinerja model. Model kemudian dilatih dengan memanggil fungsi `fit(X_train_tfidf, y_train)` untuk

mempelajari hubungan antara fitur teks dengan kelas target.

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB(alpha=0.1) # Coba juga alpha 0.1 atau 0.3
model.fit(X_train_tfidf, y_train)

MultinomialNB
MultinomialNB(alpha=0.1)
```

Gambar 3. Pelatihan model naïve bayes

## HASIL DAN PEMBAHASAN

### A. Evaluasi (Confusion Matrix)

Tahap evaluasi dilakukan untuk menilai kinerja model klasifikasi dalam mengidentifikasi sentimen ulasan produk. Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan label aktual dalam data uji, menggunakan metrik seperti akurasi, presisi, recall, skor F1, dan matriks kebingungan untuk analisis.

Hasil evaluasi menunjukkan bahwa model mencapai akurasi 83,33%, dengan presisi 0,84, recall 0,83, dan skor F1 0,83. Nilai-nilai ini menunjukkan bahwa model berkinerja cukup baik dalam mengklasifikasikan tiga kategori sentimen, yaitu sangat sesuai, sesuai, dan tidak sesuai. Gambar 4 mengilustrasikan visualisasi akurasi model, yang mengungkapkan bahwa model dapat mencapai tingkat akurasi 83,33% pada data uji.

```
Akurasi: 0.8333
Precision: 0.84
Recall: 0.83
F1-Score: 0.83

Classification Report:
              precision    recall  f1-score   support

cocok         0.91     0.72     0.81     40
sangat_cocok 0.85     0.91     0.88     32
tidak_cocok   0.75     0.50     0.62     30

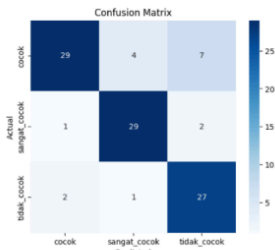
accuracy      0.84     0.84     0.83    102
macro avg     0.84     0.84     0.83    102
weighted avg  0.84     0.83     0.83    102

Confusion Matrix:
[[29  4  7]
 [ 1 29  2]
 [ 2  1 27]]
```

Gambar 4. Hasil Akurasi, Precision, Recall, F1-Score

Analisis confusion matrix menunjukkan bahwa kelas sangat sesuai dan tidak sesuai terklasifikasi dengan tingkat akurasi yang relatif tinggi. Sementara itu, pada kelas sesuai masih ditemukan sejumlah kesalahan klasifikasi, di mana beberapa ulasan salah diklasifikasikan menjadi sangat sesuai atau tidak sesuai. Hal ini menunjukkan bahwa model mengalami ambiguitas dalam membedakan ekspresi sentimen positif dengan intensitas yang sama.

Oleh karena itu, perbaikan model dapat diarahkan melalui penyesuaian parameter (*hyperparameter tuning*), pemilihan fitur yang lebih representatif, atau eksplorasi metode klasifikasi yang lebih kompleks, seperti Support Vector Machine atau ensemble learning. Hasil confusion matrix dapat dilihat pada gambar 5.



Gambar 5. Confusion Matrix

B. Visualisasi (WordCloud)

Pada tahap akhir penelitian ini dilakukan visualisasi hasil klasifikasi sentimen ulasan produk sunscreen dari pengguna platform Female Daily. Visualisasi ini berbentuk WordCloud yang bertujuan untuk menampilkan kata-kata yang paling sering muncul pada setiap kategori sentimen, yaitu sangat cocok, cocok, dan tidak cocok. Setiap WordCloud merepresentasikan karakteristik opini pengguna secara visual berdasarkan frekuensi kemunculan kata tersebut, sehingga membantu memahami pola bahasa dan

memfokuskan perhatian pengguna pada setiap tingkat kepuasan terhadap produk sunscreen.



Gambar 6. Wordcloud cocok



Gambar 7. Wordcloud sangat cocok



Gambar 8. Wordcloud tidak cocok

KESIMPULAN

Studi ini menunjukkan bahwa algoritma *Naïve Bayes* dapat digunakan secara efektif untuk mengklasifikasikan sentimen dalam ulasan produk tabir surya dari platform Female Daily. Melalui serangkaian tahapan, mulai dari pengumpulan data, pra-pemrosesan teks, pelabelan sentimen berdasarkan rating, oversampling untuk penyeimbangan data, hingga transformasi data menggunakan metode TF-IDF, model berhasil melakukan klasifikasi multikelas dengan hasil yang cukup baik. Hasil evaluasi menggunakan matriks kebingungan mengungkapkan bahwa model tersebut mencapai skor akurasi 83,33%, dengan presisi 0,84, recall 0,83, dan skor F1 0,83. Hasil ini menunjukkan bahwa model tersebut berkinerja baik dalam mengklasifikasikan setiap kategori sentimen yang dianalisis, yang meliputi sangat sesuai,

sesuai, dan tidak sesuai. Namun, masih terdapat beberapa kesalahan klasifikasi pada kelas sesuai, di mana model mengalami ambiguitas dalam membedakan intensitas sentimen positif sedang dengan kelas lainnya. Hal ini menunjukkan perlunya penyesuaian lebih lanjut agar model dapat lebih peka terhadap perbedaan intensitas opini pengguna. Visualisasi menggunakan WordCloud pada setiap kelas sentimen memperkuat hasil klasifikasi model. Kata-kata seperti ringan, cepat serap, dan nyaman mendominasi kelas positif, sedangkan kata-kata seperti jerawat, gatal, dan perih muncul pada kelas negatif, yang menunjukkan konsistensi antara label sentimen dan konten ulasan.

Berdasarkan temuan ini, beberapa langkah pengembangan dan penyempurnaan dapat diusulkan untuk meningkatkan kualitas dan akurasi model secara keseluruhan. Disarankan agar penelitian selanjutnya mempertimbangkan penggunaan teknik klasifikasi lain yang lebih kompleks seperti Support Vector Machine atau ensemble learning untuk meningkatkan akurasi klasifikasi, terutama pada kelas sentimen yang cenderung ambigu. Selain itu, eksplorasi pemilihan fitur yang lebih kontekstual dan penyetelan parameter (*hyperparameter tuning*) juga dapat membantu meningkatkan kinerja model. Penambahan data dari berbagai produk sunscreen juga dapat memperluas cakupan dan memperkuat model generalisasi. Studi ini diharapkan dapat menjadi landasan bagi kemajuan sistem pendukung keputusan berbasis opini pengguna dan membantu dalam memperoleh wawasan lebih dalam tentang persepsi konsumen dalam industri kecantikan.

## **REFERENSI**

# Klasifikasi Sentimen Ulasan Sunscreen-1.docx

## ORIGINALITY REPORT

16%

SIMILARITY INDEX

8%

INTERNET SOURCES

14%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

- 1 Nur Aisyah Wahyuni, Dinda Putri Ayu, Hafidz Irsyad. "Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine", Arcitech: Journal of Computer Science and Artificial Intelligence, 2024 2%

Publication
- 2 Awang Herjunie Nurdy, Abdul Rahim, Arbansyah. "Analisis Sentimen Ulasan Game Stumble Guys Pada Playstore Menggunakan Algoritma Naïve Bayes", Teknika, 2024 1%

Publication
- 3 Siti Aiwestopa Riyandona, Nining Rahaningsih, Raditya Dinar Dana, - Mulyawan. "IMPLEMENTASI MODEL ANALISIS SENTIMEN TERHADAP GRUP MUSIK BTS MENGGUNAKAN METODE NAÏVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2025 1%

Publication
- 4 [journal.unipdu.ac.id](http://journal.unipdu.ac.id) 1%

Internet Source
- 5 Haris Junianto, Primandani Arsi, Bagus Adhi Kusuma, Dhanar Intan Surya Saputra. "Evaluasi Aplikasi Raileo Melalui Analisis Sentimen Ulasan Playstore Dengan Metode Naive Bayes", SINTECH (Science and Information Technology) Journal, 2024 1%

Publication
- 6 [pmc.ncbi.nlm.nih.gov](http://pmc.ncbi.nlm.nih.gov)

Internet Source

1%

7

Submitted to Universitas Kristen Wira Wacana  
Sumba

Student Paper

1%

8

Rizqia Lestika Atimi, Enda Esyudha Pratama.  
"Implementasi Model Klasifikasi Sentimen  
Pada Review Produk Lazada Indonesia", Jurnal  
Sains dan Informatika, 2022

Publication

1%

9

Ilm Muhaemin Abdul Azis, Tri Wahyudi.  
"Analisa Sentimen Rencana Pemindahan Ibu  
Kota Nusantara dari Jakarta ke Kalimantan  
Timur Menggunakan Algoritma Naïve Bayes",  
Jurnal Indonesia : Manajemen Informatika  
dan Komunikasi, 2024

Publication

1%

10

Ilham Fariz Asya Mubarok, Baenil Huda,  
Agustia Hananto, Tukino Tukino, Huban Kabir.  
"ANALISIS USER SENTIMENT APLIKASI  
GOOGLE MAPS, MAPS.ME DAN WAZE  
MENGUNAKAN METODE SUPPORT VECTOR  
MACHINE", Rabit : Jurnal Teknologi dan Sistem  
Informasi Univrab, 2023

Publication

1%

11

frangao.net

Internet Source

1%

12

Submitted to UIN Sultan Syarif Kasim Riau

Student Paper

1%

13

Rizki Rino Pratama. "Analisis Model Machine  
Learning Terhadap Pengenalan Aktifitas  
Manusia", MATRIK : Jurnal Manajemen, Teknik  
Informatika dan Rekayasa Komputer, 2020

Publication

1%

14	Internet Source	1 %
15	shopee.co.id Internet Source	<1 %
16	www.scribd.com Internet Source	<1 %
17	Christ Mario, Ryan Randy Suryono. "Public Sentiment Analysis on Dirty Vote Movie on YouTube using Random Forest and Naïve Bayes", INOVTEK Polbeng - Seri Informatika, 2024 Publication	<1 %
18	escholarship.org Internet Source	<1 %
19	id.123dok.com Internet Source	<1 %
20	riaulink.com Internet Source	<1 %
21	msp1120elvissimarmata.blogspot.com Internet Source	<1 %
22	repository.uinsu.ac.id Internet Source	<1 %
23	Anindita Puspa Ayu Prayogi, Altha Inas Shofyana, Dewi Putriani. "Prediksi <i>Credit Card Approval</i> Menggunakan Algoritma <i>Random Forest</i> ", Jurnal Ilmu Komputer dan Multimedia, 2025 Publication	<1 %
24	Muhammad Rizky Julianto, Yuma Akbar, Tri Wahyudi. "Analisis Sentimen Respon Publik Terhadap Program Internet Gratis di Platform X Melalui Pendekatan Algoritma Naïve Bayes", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2024	<1 %

25

Aang Alim Murtopo, Maulana Aditdya, Pingky Septiana Ananda, Gunawan Gunawan. "PENERAPAN COMPUTER VISION UNTUK MENDETEKSI KELENGKAPAN ATRIBUT SISWA MENGGUNAKAN METODE CNN", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2024

<1%

Publication

---

26

Romadon Goring Siregar, Aidil Halim Lubis, Muhammad Ikhsan. "Klasifikasi Pengaruh Negatif Game online Bagi Remaja Menggunakan Algoritma Naïve Bayes", DEVICE : JOURNAL OF INFORMATION SYSTEM, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, 2025

<1%

Publication

---

27

Siti Mariam, Ida Nurhaida. "Analisis Sentimen berbasis Deep Learning Terhadap Kesetaraan Gender di Bidang STEM: Perspektif dan Implikasinya", Edumatic: Jurnal Pendidikan Informatika, 2025

<1%

Publication

---

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off