

# ANALISIS SENTIMEN ULASAN PRODUK DAVIENA DI SHOPEE MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Faisal Reza Pradhana<sup>1</sup>, Aziz Musthafa\*<sup>2</sup>, Indah Fitria<sup>3</sup>

<sup>12</sup>universitas Darussalam Gontor, <sup>3</sup>Universitas Darussalam Gontor

<sup>12</sup>Indonesia, <sup>3</sup>Indonesia

Email: [1faisalrezapradhana@unida.gontor.ac.id](mailto:1faisalrezapradhana@unida.gontor.ac.id), [2aziz@unida.gontor.ac.id](mailto:2aziz@unida.gontor.ac.id),  
[3indahfitria42035@mhs.unida.gontor.ac.id](mailto:3indahfitria42035@mhs.unida.gontor.ac.id)

## Abstract

*The e-commerce industry in Indonesia is driving the growth of marketplaces that make online buying and selling platforms, such as Shopee. One of the most popular product categories is skincare, which is in high demand by Indonesian consumers, especially women. Local skincare brands such as Daviena have attracted attention for offering products that are able to fulfill skincare needs. However, in choosing skincare, consumers are faced with selectivity as it suits the needs of their respective skin types. Reviews of a product that include star ratings and review comments from buyers are important to evaluate a product's quality from a user's perspective. This study collected 4,000 Daviena product reviews from shopee, utilizing web scrapy to manage the data. Sentiment analysis was conducted to classify reviews into positive, neutral, and negative categories, and help product owners understand consumer perceptions. With the classification, consumers can gain a better understanding of the quality of a product. The test results show that the highest accuracy achieved is 94% with a combination of training and test data division of 90:10 using the SVM method to perform classification.*

**Keywords:** Shopee, skincare, Daviena brand, sentiment analysis.

## Abstraksi

*Industri e-commerce di Indonesia mendorong pertumbuhan marketplace yang menjadikan wadah jual beli online, seperti Shopee. Salah satu kategori produk yang paling populer ialah skincare, yang sangat diminati oleh konsumen Indonesia, terutama kaum wanita. Merk-merk skincare lokal seperti Daviena telah menarik perhatian karena menawarkan produk yang mampu memenuhi kebutuhan perawatan kulit. Namun, dalam pemilihan skincare, konsumen dihadapi dengan selektivitas karena sesuai dengan kebutuhan tipe kulit masing-masing. Ulasan sebuah produk yang meliputi rating bintang dan ulasan komentar dari pembeli menjadi penting untuk mengevaluasi sebuah kualitas produk dari perspektif pengguna. Penelitian ini mengumpulkan 4.000 ulasan produk Daviena dari shopee, dengan memanfaatkan web scrapy untuk mengelola data. Sentimen analisis dilakukan untuk mengklasifikasi ulasan menjadi kategori positif, netral, dan negatif, serta membantu pemilik produk dalam memahami persepsi konsumen. Dengan pengklasifikasian, konsumen dapat memperoleh pemahaman yang lebih baik mengenai kualitas sebuah produk. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang dicapai ialah sebesar 94% dengan kombinasi pembagian data latih dan data uji 90:10 menggunakan metode SVM dalam melakukan klasifikasi dengan baik.*

**Kata Kunci:** Shopee, skincare, Daviena, sentimen analisis.

## 1. PENDAHULUAN

Produk *skincare* semakin populer, terutama di kalangan generasi muda yang mengutamakan kemudahan dalam berbelanja melalui platform *e-commerce* seperti Shopee[1]. Ulasan dari konsumen (review) menjadi faktor krusial dalam keputusan pembelian, di mana *electronic Word Of Mouth* (e-WOM) dan *rating* terbukti memiliki pengaruh yang signifikan terhadap preferensi konsumen[2]. Analisis ulasan *online* memberikan wawasan penting mengenai permintaan konsumen dan dapat membantu pedagang dalam merancang strategi pemasaran yang lebih efektif.

Namun volume ulasan yang besar dengan format yang tidak terstruktur serta adanya berbagai sentimen (positif, negatif, netral) sehingga membuat analisis manual menjadi tantangan[3][4]. Teknik *text mining* pada ulasan konsumen memungkinkan pengelompokan dan ekstraksi fitur dari teks ulasan, sehingga dapat menghasilkan informasi yang bernilai bisnis. Hal ini membantu pedagang dan platform *e-commerce* untuk merancang strategi peningkatan layanan yang relevan, mendorong pertumbuhan bisnis, serta meningkatkan kepuasan pelanggan.

Analisis sentimen adalah sub-bidang dari pemrosesan bahasa alami (Natural Language Processing atau NLP) yang bertujuan untuk mengklasifikasikan sentimen dari teks bebas secara otomatis, dan telah diterapkan secara luas untuk memahami opini konsumen di platform *e-commerce* dan media sosial[5]. Metode *machine learning* yang sering digunakan meliputi Naive Bayes[6], Support Vector Machine (SVM) [5], dan metode *deep learning* seperti Long Short-Term Memory (LSTM) serta Convolutional Neural Networks (CNN) [3]. Naive Bayes dikenal karena kesederhanaan dan efisiensinya, sementara SVM sering dipilih karena kemampuannya dalam menangani data yang tidak seimbang dan memberikan akurasi tinggi. Di sisi lain, metode *deep learning* semakin populer karena kemampuannya menangkap hubungan kompleks dalam data teks dan menghasilkan prediksi yang lebih akurat dalam beberapa kasus.

Namun, analisis sentimen masih menghadapi tantangan, terutama dalam menangani teks yang tidak terstruktur dan informal seperti ulasan produk dan komentar di media sosial. Kurangnya korpus dan leksikon khusus domain sering kali menurunkan akurasi model[7]. Kesalahan tata bahasa dan penggunaan istilah yang tidak dikenal juga memengaruhi hasil analisis. Untuk mengatasi masalah ini, penelitian di masa depan diharapkan dapat memanfaatkan teknik *praproses* data yang lebih canggih, penggunaan model *deep learning* yang lebih kompleks, serta penerapan metode pembelajaran semi-terawasi atau tidak terawasi guna meningkatkan kinerja analisis sentimen[3].

Penelitian [5] menunjukkan bahwa Support Vector Machine (SVM) efektif dalam klasifikasi teks ulasan produk dengan akurasi mencapai 89.98%. [8] juga mengonfirmasi keunggulan SVM, mencatat akurasi lebih dari 80% dalam menganalisis sentimen terhadap produk *skincare*. Sementara itu, [9] mengeksplorasi penggunaan Word2Vec sebagai fitur dalam analisis sentimen dengan SVM, meski hasilnya masih lebih rendah dibandingkan metode lain. Beberapa penelitian telah membandingkan SVM dengan Naive Bayes dalam konteks *e-commerce*, seperti pada produk *earphone* di Tokopedia [6], dan pada ulasan

aplikasi Shopee yang mencapai akurasi 84.71% [10]. Teknik SMOTE juga digunakan dalam kombinasi dengan SVM, yang meningkatkan akurasi hingga 92% [11]. Studi terkait yang memanfaatkan analisis sentimen dari ulasan YouTube untuk pengembangan produk baru menunjukkan pendekatan berbeda dibandingkan penelitian pada ulasan e-commerce [12].

Dari penelitian-penelitian tersebut, terlihat bahwa meskipun banyak studi telah dilakukan menggunakan metode SVM dalam konteks analisis sentimen, masih terdapat kekurangan dalam hal penerapan spesifik pada kategori produk tertentu seperti *skincare* Daviena dan platform *e-commerce* lokal seperti Shopee. Penelitian ini bertujuan untuk menutup celah tersebut dengan memberikan analisis mendalam tentang sentimen konsumen terhadap produk *skincare* Daviena melalui pendekatan yang terfokus.

Tujuan utama dari penelitian ini adalah untuk menerapkan metode Support Vector Machine (SVM) dalam menganalisis sentimen ulasan produk *skincare* Daviena di platform *e-commerce* Shopee. Diharapkan hasil penelitian ini dapat memberikan wawasan yang berguna bagi produsen dan pemasar dalam memahami pandangan konsumen serta meningkatkan kualitas produk dan layanan mereka berdasarkan umpan balik yang diperoleh dari ulasan tersebut.

## 2. TINJAUAN PUSTAKA

Tinjauan pustaka bertujuan untuk mengulas studi terdahulu yang relevan dengan topik, metode, dan hasil penelitian yang akan dilakukan. Dalam bagian ini, pendekatan, hasil utama, kelebihan, dan kekurangan dari penelitian sebelumnya akan dirinci untuk mengidentifikasi peluang penelitian baru dan memperkuat dasar teoritis. Bagian ini juga menyoroti gap penelitian yang menjadi fokus utama penelitian ini.

### 2.1. Analisis Sentimen

Analisis sentimen telah menjadi alat penting dalam memahami opini publik terhadap produk, layanan, atau topik tertentu [13][14]. [5] menunjukkan bahwa algoritma Support Vector Machine (SVM) efektif dalam klasifikasi ulasan produk dengan akurasi mencapai 89.98%. Studi ini menyoroti keunggulan SVM dalam menangani data teks tidak terstruktur, terutama dalam tugas-tugas klasifikasi sentimen. Namun, penelitian tersebut terbatas pada fitur tradisional seperti *Bag of Words*, yang dapat kehilangan konteks semantik dari ulasan. Sementara itu, [8] fokus pada ulasan produk *skincare* dengan hasil akurasi lebih dari 80%, tetapi penelitian ini tidak mengeksplorasi metode optimalisasi fitur atau teknik *balancing dataset*, yang berpotensi meningkatkan akurasi lebih jauh.

### 2.2. Penggunaan Fitur dalam Analisis Sentimen

Dalam analisis sentimen, pemilihan fitur memainkan peran penting. [9] mengeksplorasi penggunaan Word2Vec sebagai representasi fitur dalam analisis sentimen berbasis SVM. Meski pendekatan ini lebih baik dalam menangkap hubungan semantik antar kata, akurasi model masih lebih rendah dibandingkan dengan metode lain

seperti TF-IDF [9]. Hal ini menunjukkan bahwa pemilihan fitur yang sesuai harus disesuaikan dengan karakteristik dataset dan algoritma yang digunakan.

### 2.3. Perbandingan Algoritma

Beberapa penelitian telah membandingkan SVM dengan algoritma lain dalam analisis sentimen. Penelitian terkait ulasan aplikasi Shopee oleh [10] menunjukkan bahwa SVM memberikan akurasi 84.71%, mengungguli algoritma Naïve Bayes yang hanya mencapai 78%. Studi ini mengonfirmasi keunggulan SVM dalam menangani data teks yang kompleks, meskipun tidak mencakup teknik pra-pemrosesan data yang mendalam seperti *stemming* atau *stopword removal*, yang dapat memengaruhi hasil akhir.

### 2.4. Optimalisasi Model

Untuk meningkatkan performa model pada *dataset* yang tidak seimbang, Teknik SMOTE (Synthetic Minority Oversampling Technique) sering digunakan dalam kombinasi dengan SVM. Ardiansyah et al. (2020) menunjukkan bahwa penggunaan SMOTE dapat meningkatkan akurasi model hingga 92%, terutama pada *dataset* yang memiliki distribusi kelas yang timpang [11]. Meski demikian, penelitian ini hanya fokus pada teknik *balancing* tanpa membahas dampaknya terhadap waktu komputasi.

### 2.5. Studi Terkait di E-Commerce

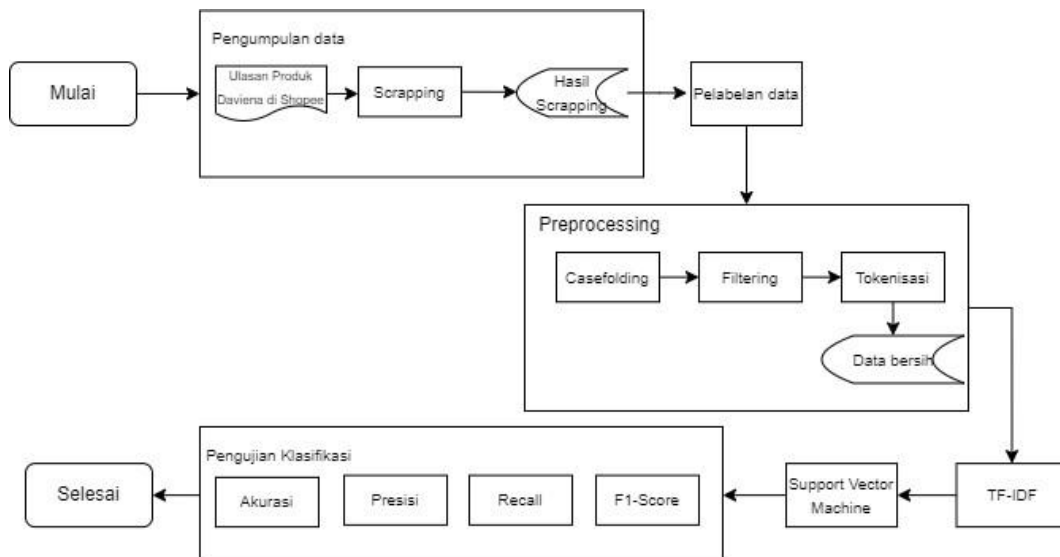
Penelitian lain mengangkat sentimen ulasan pada platform *e-commerce*. [5] dan [9] telah mengkaji ulasan produk di Tokopedia dan YouTube, masing-masing, dengan fokus pada pengembangan produk baru dan analisis persepsi konsumen. Studi ini menawarkan wawasan yang luas mengenai penggunaan algoritma *machine learning* dalam berbagai domain. Namun, ulasan pada platform Shopee, khususnya terkait produk *skincare* lokal seperti Daviena, masih belum banyak dieksplorasi, membuka peluang untuk penelitian lebih lanjut.

Dari berbagai literatur yang dibahas, terdapat beberapa kekurangan dan celah penelitian yang dapat diisi oleh studi ini: (1) Kurangnya penelitian yang secara khusus menganalisis ulasan produk *skincare* lokal seperti Daviena pada platform Shopee. (2) Belum banyak penelitian yang menggabungkan SVM dengan optimalisasi fitur dan teknik *balancing dataset* seperti SMOTE untuk meningkatkan akurasi. (3) Keterbatasan eksplorasi terhadap dampak metode *preprocessing* data yang berbeda, seperti *stemming* dan *lemmatization*, terhadap hasil analisis sentimen.

Berdasarkan tinjauan pustaka ini, penelitian ini bertujuan untuk mengisi gap tersebut dengan menganalisis sentimen ulasan produk *skincare* Daviena di Shopee menggunakan algoritma SVM, sekaligus menguji efektivitas kombinasi metode *preprocessing*, fitur TF-IDF, dan teknik *balancing dataset*. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam konteks analisis sentimen *e-commerce*, khususnya pada produk lokal di Indonesia.

### 3. METODE PENELITIAN

Proses penelitian ini mengacu pada pengumpulan data, analisis dan evaluasi model, serta melihat performa klasifikasi berdasarkan data ulasan konsumen. Penelitian ini menggunakan data ulasan produk oleh konsumen yang diambil dari shopee menggunakan scraping. Proses Alur penelitian ditampilkan pada gambar 1.



Gambar 1. Metodologi Penelitian[15]

Berikut tahapan yang dilakukan:

#### Pengumpulan dataset

Penelitian ini menggunakan data ulasan produk perawatan dari shopee. Data ini diperoleh dengan menggunakan teknik scraping dengan berbasis aplikasi scaper. Data yang diambil berjumlah 4000 data ulasan. Data tersebut disimpan dalam bentuk csv.

Setelah proses pengumpulan data, tahap selanjutnya ialah pelabelan data atau pemberian tag pada data. Pemberian tag mencakup sentiment positif, negatif, ataupun netral. Data yang tergolong sentiment positif berupa pujian, apresiasi, atau ungkapan kepuasan terhadap produk tersebut. Dan jika yang tergolong sentiment negatif berupa kritik, ketidakpuasan, atau masalah yang dihadapi setelah penggunaan produk. Sedangkan jika ulasan netral berupa ketidakberpihakan untuk menilai produk secara negatif ataupun positif.

#### Preprocessing

Tahap membersihkan kata-kata yang tidak diperlukan. Pada proses ini seluruhnya menggunakan bahasa python. Preprocessing yang dilakukan diantaranya ialah:

##### Casefolding

perubahan semua karakter huruf pada review menjadi huruf kecil semua

##### Filtering

berfungsi untuk menghilangkan karakter khusus, seperti \$,%,#, % dan lain sebagainya. Filtering juga berfungsi menghilangkan kata yang tidak sesuai dengan hasil parsing seperti hastag, #, username, @, emoticon, dan URL. Dan untuk simbol, tanda atau angka juga akan dihapus karena tidak memiliki banyak pengaruh pelabelan khusus.

Tokenizing

ialah membagi ulasan dari kalimat menjadi perkata.

TF-IDF

Metode TF-IDF ialah hasil perkalian antara Term Frequency (TF) dan Inverse Document Frequency (IDF). Yang memberikan bobot kepada kata-kata yang sering muncul dalam suatu dokumen tapi jarang muncul didokumen lain, sehingga memperhitungkan kepentingan data tersebut dalam konteks dokumen tertentu. Term Frequency (TF) adalah ukuran seberapa sering kata muncul dalam sebuah dokumen. Inverse Document Frequency (IDF) adalah ukuran seberapa jarang sebuah kata muncul diseluruh dokumen dalam korpus teks.

Klasifikasi menggunakan metode SVM

Digunakan splitting data yang dibagi menjadi dua bagian , yaitu data uji dan data latih. Penulis melakukan 3 percobaan pada penelitian ini.

Percobaan pertama, 90% data digunakan sebagai data latih dan 10% sebagai data uji

Percobaan kedua, 80% data sebagai data latih dan 20% data uji

Percobaan ketiga, 70% data sebagai data latih dan 30% data uji

Pengujian klasifikasi

Setelah proses klasifikasi maka selanjutnya akan dilakukan pengujian klasifikasi dengan menggunakan confusion matrix guna mengetahui kinerja model. Evaluasi dilakukan untuk mengetahui tingkat akurasi yang didapatkan model klasifikasi saat pemrosesan data. Setelah mengetahui hasil, maka dapat dilihat bagaimana metode klasifikasi bekerja pada setiap kelasnya melalui nilai tersebut yaitu, akurasi, presisi, recall, dan f1-score. Masing-masing memiliki nilai 0-1. Semakin tinggi nilainya maka semakin baik.

#### **4. HASIL DAN PEMBAHASAN**

Tahapan yang dilakuakn pada tahap implementasi penelitian yang dilakukan sebagai berikut.

##### **4.1. Import Library**

Sebelum dilakukan proses analisa maka dilakukan pengimporan library yang dibutuhkan dalam proses analisis data. Pada gambar 2 disajikan dengan kode pemrograman python.

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.feature_extraction.text import TfidfVectorizer
6 from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score
7 from sklearn.metrics import classification_report
8 import nltk
9
10 nltk.download('punkt')
11 from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer, HashingVectorizer
12 from sklearn.pipeline import Pipeline
13 from wordcloud import WordCloud, STOPWORDS
14 import re
```

Gambar 2. Import Library

## 4.2. Memuat data

Tahapan pembacaan dataset yang telah didapatkan sebelumnya. Pada gambar 3 disajikan dengan kode pemrograman python.

```
1 #memuat data
2 df = pd.read_csv('Documents/data/dataScraping_Daviena.csv', index_col=0)
3 #menampilkan data paling atas
4 df
```

Gambar 3. Memuat Data

Setelah program dijalankan, maka akan menampilkan 10 baris data. Hasil seperti pada gambar 4 dibawah ini.

	Nama Pelanggan	Produk	Ulasan	Rating
0	S***a	Body Lotion Daviena Skincare Power of Moonlight	baru pertama kali coba, mudah mudahan cocok	5
1	e***a	Body Lotion Daviena Skincare Power of Moonlight	bagus	4
2	P***a	Body Lotion Daviena Skincare Power of Moonlight	Ok pengiriman cepat. nitip foto beforenya ya d...	5
3	W***a	Body Lotion Daviena Skincare Power of Moonlight	semoga si kuning ini bs membuat kulitku menjad...	5
4	L***a	Body Lotion Daviena Skincare Power of Moonlight	Baru mau mencoba semoga cocok dan hasilnya bis...	5
...	...	...	...	...
4064	C***a	Acne Gel Strawberry Daviena Skincare Acne Seri...	pengiriman cepat, smoga cocok, thanks seller	5
4065	Yogi	Acne Gel Strawberry Daviena Skincare Acne Seri...	mantapp, sejak pake ini jadi jarang jerawat...	5
4066	George	Acne Gel Strawberry Daviena Skincare Acne Seri...	exp cepat amat	3
4067	Sanny	Acne Gel Strawberry Daviena Skincare Acne Seri...	Packing baik, aman. Mudah2an cocok toner ya. T...	5
4068	Dania	Acne Gel Strawberry Daviena Skincare Acne Seri...	thank you produknya sudah sampe dengan selamat...	5

Gambar 4. Hasil Pemuatan Dataset

## 4.3. Cleaning data

```
1 #hanya memilih kolom produk, ulasan, rating
2 df = df[['Produk', 'Ulasan', 'Rating']]
3
1 df.shape
(4069, 3)
1 df = df.drop_duplicates(subset='Ulasan')
1 df = df.dropna()
1 df.shape
(3086, 3)
1 df
```

Gambar 5. Cleaning Dataset

Dilakukan cleaning dataset pada gambar 5, yang hanya memilih kolom produk, ulasan, rating saja, selain kolom yang dipilih maka akan dihapus. Diperoleh hasil dari pemrograman disajikan pada gambar 6 dibawah ini.

	Produk	Ulasan	Rating	label
0	Body Lotion Daviena Skincare Power of Moonlight	baru pertana kali coba mudah mudahan cocok	5	positif
1	Body Lotion Daviena Skincare Power of Moonlight	pengiriman cepat nitip foto befornya disini ne...	5	positif
2	Body Lotion Daviena Skincare Power of Moonlight	semoga si kuning bs membuat kulitku menjadi se...	5	positif
3	Body Lotion Daviena Skincare Power of Moonlight	baru mau mencoba semoga cocok hasilnya sesuai ...	5	positif
4	Body Lotion Daviena Skincare Power of Moonlight	dicobasemoga cocok	5	positif
...	...	...	...	...
2964	Acne Gel Strawberry Daviena Skincare Acne Seri...	pengiriman cepat smoga cocok thanks seller	5	positif
2965	Acne Gel Strawberry Daviena Skincare Acne Seri...	mantapp sejak pake jadi jarang jerawat	5	positif
2966	Acne Gel Strawberry Daviena Skincare Acne Seri...	exp cepat	3	netral
2967	Acne Gel Strawberry Daviena Skincare Acne Seri...	packing baik aman mudah2an cocok toner trimakasih	5	positif
2968	Acne Gel Strawberry Daviena Skincare Acne Seri...	thank you produknya sampe selamat gak bocor pa...	5	positif

Gambar 6. Hasil Cleaning Dataset

#### 4.4. Labeling

Pelabelan dilakukan untuk mengkategorikan data kedalam kelompok positif, netral, dan negatif. Untuk label positif memberikan rating 4 dan 5, untuk label netral dengan rating 3, dan untuk label negatif dengan rating 1 dan 2. Dapat dilihat pada gambar 7 dibawah ini.

```

1 label = []
2 for index, row in df.iterrows():
3     if row["Rating"] == 5 or row["Rating"] == 4:
4         label.append('positif')
5     elif row["Rating"] == 3:
6         label.append('netral')
7     else:
8         label.append('negatif')
9
10 df["label"] = label
    
```

Gambar 7. Labelling Dataset

Berikut ialah hasil dari proses labeling, jika ulasan berbentuk kalimat pujian maka dilabeli positif, jika ulasan berbentuk kalimat kritik maka dilabeli negatif. Dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Hasil Proses Labeling

No	Ulasan	Label
1.	cocoke pakai produk	Positif
2.	tidak ngefek di aku	Netral
3.	krim malemnya terlalu lengket gak enak di pakainya	Negatif

#### 4.5. Preprocessing

Data yang sudah melewati labelling selanjutnya akan melewati proses preprocessing terlebih dahulu. Proses preprocessing ini sangat membantu membersihkan dan menyiapkan teks sehingga lebih mudah untuk dianalisis oleh algoritma machine learning. Proses preprocessing pada sentiment analisis tidak boleh dilewatkan, karena tanpa preprocessing yang tepat, maka hasil analisis dapat menjadi tidak akurat. Terdapat beberapa tahapan yaitu:

##### 4.5.1. Casefolding/Lowercasing

Pada tahapan pertama preprocessing ialah case folding. Pada proses ini bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil. Hasil casefolding dapat dilihat pada tabel 2 dibawah ini.



Tabel 2. Hasil Casefolding

Sebelum CaseFolding	Setelah Casefolding
barang ORI pengiriman cepat	barang ori pengiriman cepat
Wangi banget, enak ! Semoga cocok dan cepet nyerahinnya	wangi banget enak semoga cocok dan cepet nyerahinnya
Bismillah baru mau coba produk daviena, semoga cocok.	bismillah baru mau coba produk daviena, semoga cocok
REVIEW JUJUR SELLER BINTANG 5 PRODUK BINTANG 1	review jujurseller bintang 5 produk bintang 1
Enak dipakai diwajah semoga berkurang bolong2nya	Enak dipakai diwajah semoga berkurang bolong2nya

#### 4.5.2. Filtering (stopword)

Selanjutnya akan dilakukan penghapusan stopwords sebagai salah satu langkah untuk memebersihkan data. Stopword adalah kata- kata yang sering muncul dalam teks, seperti “ dan”, “ tidak”, “yang”, “untuk”. Hasil stopwords dapat dilihat pada tabel 3 dibawah ini.

Tabel 3. Hasil Filtering

Sebelum	Sesudah
Baru mau mencoba semoga cocok dan hasilnya bisa sesuai ekspektasi ya.	baru mau mencoba semoga cocok hasilnya sesuai ekspektasi.

#### 4.5.3. Tokenizing

Proses Tokenizing ialah proses Dimana teks diubah menjadi daftar kata-kata individual, yang kemudian dapat dianalisis lebih lanjut dalam sentiment analisis. Hasil tokenizing dapat dilihat pada tabel 4 dibawah ini.

Tabel 4. Hasil Tokenizing

Sebelum	Sesudah
semoga si kuning ini bs membuat kulitku menjadi sebening bihun	semoga, si, kuning, bisa, buat, kulit, jadi, bening, bihun.

#### 4.6. Pembobotan TF-IDF

Setelah melewati proses preprocessing maka data akan memasuki proses pembobotan. Pembobotan pada penelitian ini menggunakan TF-IDF . TF-IDF bekerja dengan menghitung seberapa sering sebuah kata muncul dalam sebuah dokumen dan setiap kata yang muncul akan di beri bobot, sehingga model bisa lebih mudah mengenali sentiment dari teks. Semakin sering sebuah kata muncul dalam suatu dokumen maka semakin tinggi nilainya. Setelah penerapan TF-IDF pada teks. Hasil dari pembobotan TF-IDF dapat dilihat pada gambar 8 dibawah ini. Baris pertama pada data yang telah dilakukan pembobotan yaitu, (0, 3731) 0.4184305629986757 dapat dijelaskan bahwa angka 0 pada kolom pertama adalah jumlah dokumen, dan pada kolom kedua ialah letak katanya, dan untuk kolom ketiga ialah hasil dari tf.

```
[8.34180784 8.34180784 8.34180784 ... 8.34180784 8.34180784 8.34180784]
(0, 3731) 0.4184305629986757
(0, 3000) 0.15432078306806501
(0, 2838) 0.24081371647921193
(0, 2727) 0.4184305629986757
(0, 2671) 0.3110834379631301
(0, 1592) 0.2540468297231768
(0, 1418) 0.32082243123077114
(0, 1208) 0.25901729451416133
(0, 821) 0.1500968473983017
(0, 485) 0.4184305629986757
(0, 185) 0.1976934237905691
```

Gambar 8. Hasil Pembobotan TF-IDF

#### 4.7. Implementasi Model SVM

Setelah melewati proses preprocessing dan pembobotan, selanjutnya data akan dilakukan processing dengan menggunakan metode Support Vector Machine. Dapat dilihat pada kode pemrograman python dibawah ini.

```
1 from sklearn.model_selection import train_test_split
2 from sklearn.svm import SVC
3 from sklearn.metrics import classification_report, accuracy_score

1 svm_model = SVC(kernel='linear')
2 svm_model.fit(X_train, y_train)

SVC(kernel='linear')
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

1 y_pred = svm_model.predict(X_test)

1 print("Accuracy:", accuracy_score(y_test, y_pred))
2 print("Classification Report:")
3 print(classification_report(y_test, y_pred))
```

Gambar 9. Model SVM

#### 4.8. Pengujian

Pengujian dengan confusion matrix akan menghasilkan nilai accuration, precision, recall, dan f-measure. Pada gambar 10 dan 11 disajikan kode progam python dari proses pengujian confusion matrix.

```
: 1 y_pred = svm_model.predict(X_test)

: 1 print("Accuracy:", accuracy_score(y_test, y_pred))
2 print("Classification Report:")
3 print(classification_report(y_test, y_pred))
```

Gambar 10. Pengujian Model SVM dengan Confusion Matrix

```
1 from sklearn.metrics import classification_report
2
3 pred = svm_model.predict(X_test)
4 print(classification_report(y_test, pred))
```

Gambar 11. Pengujian Model SVM dengan Confusion Matrix

Proses pengujian data yang dilakukan menggunakan beberapa kombinasi yaitu 90:10, 80:20, dan 70:30. Maka diperoleh hasil akurasi tertinggi ialah 94% dengan kombinasi data uji 90:10. Hasil dapat dilihat pada gambar 12 dibawah ini dan gambar 13 menunjukkan hasil data uji dengan model grafik.

```

=====
-----Data Training 90%-----
precision  recall  f1-score  support

negatif    0.00    0.00    0.00     4
netral     0.00    0.00    0.00     5
positif    0.97    1.00    0.99    300

accuracy   0.97    0.97    0.97    309
macro avg  0.32    0.33    0.33    309
weighted avg 0.94    0.97    0.96    309
=====
-----Data Training 80%-----
precision  recall  f1-score  support

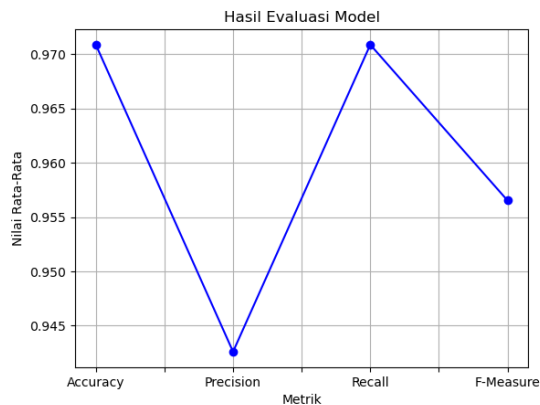
negatif    0.00    0.00    0.00     8
netral     0.00    0.00    0.00    10
positif    0.97    1.00    0.99    600

accuracy   0.97    0.97    0.97    618
macro avg  0.32    0.33    0.33    618
weighted avg 0.94    0.97    0.96    618
=====
-----Data Training 70%-----
precision  recall  f1-score  support

negatif    0.00    0.00    0.00    12
netral     0.00    0.00    0.00    14
positif    0.97    1.00    0.99    900

accuracy   0.97    0.97    0.97    926
macro avg  0.32    0.33    0.33    926
weighted avg 0.94    0.97    0.96    926
=====
    
```

Gambar 12. Hasil Pengujian SVM dengan Confusion Matrix



Gambar 13. Hasil Evaluasi Model

## 5. KESIMPULAN

Hasil penelitian yang diperoleh ialah Metode SVM yang digunakan pada ulasan produk Daviena menunjukkan akurasi tertinggi dengan 94% dengan kombinasi data latihan dan data uji 90:10. SVM dapat menunjukkan secara efektif hasil pengklasifikasian ulasan menjadi positif, netral dan negatif. Evaluasi menggunakan matrik akurasi, presisi, recall, dan f1-score yang menunjukkan kinerja model yang baik.

Adapun kelebihan, ialah: Metode SVM memiliki keakuratan yang tinggi dalam mengolah data dengan jumlah besar, seperti data ulasan yang diperoleh dari shopee. TF-IDF digunakan sebagai pembobotan kata memungkinkan pemodelan yang lebih relevan, dengan memberi bobot lebih tinggi pada kata-kata yang penting. Proses preprocessing berhasil meningkatkan kualitas data sehingga menghasilkan data yang lebih akurat.

Adapun kekurangannya, ialah: Penggunaan data ulasan dari shopee membatasi generalisasi hasil penelitian ini ke platform lain yang memiliki format dan gaya ulasan yang

berbeda; SVM hanya digunakan untuk mengklasifikasikan ulasan kedalam kategori positif dan negatif, sementara sentimen atau emosi yang lebih kompleks tidak terdeteksi.

Penelitian ini agar dapat dikembangkan lebih lanjut dengan menggunakan berbagai algoritma lainnya. Selain itu, penelitian ini dapat diperluas dengan mengumpulkan data ulasan dari berbagai platform. Serta pengembangan lainnya termasuk meningkatkan tahap preprocessing, seperti penanganan kata-kata slang yang banyak muncul dalam ulasan online.

### **DAFTAR PUSTAKA**

- [1] P. M. Putri and R. . Marlien, "Pengaruh Digital Marketing terhadap Keputusan Pembelian Online," *Jesya (Jurnal Ekon. Ekon. Syariah)*, vol. 5, no. 1, pp. 25–36, Jan. 2022, doi: 10.36778/jesya.v5i1.510.
- [2] P. W. Ningsih, C. S. Savitri, and S. P. Faddila, "Pengaruh Electronic Word Of Mouth (E-Wom) Dan Rating Terhadap Keputusan Pembelian Produk Skincare Bening's Pada Marketplace Shopee Di Kalangan Mahasiswa Kab. Karawang," *J. Econ. Bussines Account.*, vol. 7, no. 1, pp. 1592–1607, Sep. 2023, doi: 10.31539/costing.v7i1.7388.
- [3] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment Analysis of Consumer Reviews Using Deep Learning," *Sustainability*, vol. 14, no. 17, p. 10844, Aug. 2022, doi: 10.3390/su141710844.
- [4] D. E. Malla Avila, "ANALISIS SENTIMEN ULASAN PRODUK SERUM WAJAH PADA BEUTY BRAND SOMETHINC MENGGUNAKAN NAIVE BAYES CLASSIFIER," *J. Mhs. Tek. Inform.*, no. 8.5.2017, pp. 2003–2005, 2022.
- [5] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques," 2019, pp. 639–647. doi: 10.1007/978-981-13-0617-4\_61.
- [6] F. Adel Ramadhan, R. R. Permana Ruslan, and A. Zahra, "Sentiment Analysis Of E-Commerce Product Reviews For Content Interaction Using Machine Learning," *Cakrawala Repos. IMWI*, vol. 6, no. 1, pp. 207–220, Feb. 2023, doi: 10.52851/cakrawala.v6i1.219.
- [7] A. Zunic, P. Corcoran, and I. Spasic, "Sentiment analysis in health and well-being: Systematic review," *JMIR Med. Informatics*, vol. 8, no. 1, p. e16023, Jan. 2020, doi: 10.2196/16023.
- [8] V. B. Lestari and D. Amalia, "Support Vector Machine for Sentiment Analysis of PT. Paragon Technology and Innovation (Case Study of Brand Make Over and Emina Product Users on Female Daily Page – Beauty Review)," *J. Appl. Stat. Data Sci.*, vol. 1, no. 1, pp. 9–19, Mar. 2024, doi: 10.21776/ub.jasds.2024.001.01.2.
- [9] M. A. Fauzi, "Word2Vec model for sentiment analysis of product reviews in Indonesian language," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, p. 525, Feb. 2019, doi: 10.11591/ijece.v9i1.pp525-530.
- [10] S. Suswadi and M. Erkamim, "Sentiment Analysis of Shopee App Reviews Using Random Forest and Support Vector Machine," *Ilk. J. Ilm.*, vol. 15, no. 3, pp. 427–435, Dec. 2023, doi: 10.33096/ilkom.v15i3.1610.427-435.
- [11] D. I. Sumantiawan, J. E. Suseno, and W. A. Syafei, "Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction," *J. Sist. Info. Bisnis*, vol. 13, no. 1, pp. 1–9, Jun. 2023, doi: 10.21456/vol13iss1pp1-9.

- [12] R. P. Limbong, Ronsen Purba, and Muhammad Fermi Pasha, "Pemanfaatan Analisis Sentimen dari Ulasan Produk di Youtube untuk Pengembangan Produk Baru," *Syntax Lit. ; J. Ilm. Indones.*, vol. 9, no. 7, pp. 3796–3810, Jul. 2024, doi: 10.36418/syntax-literate.v9i7.13568.
- [13] A. Hanafiah, A. H. Nasution, Y. Arta, R. Wandri, H. O. Nasution, and J. Mardafora, "Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 230–236, 2023, doi: 10.31539/intecom.v6i1.5845.
- [14] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [15] R. W. Pratiwi, S. F. H, D. Dairoh, D. I. Af'idah, Q. R. A, and A. G. F, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387.