

## Analisis Sentimen Ulasan Produk Sayur di Tokopedia Menggunakan Model Support Vector Machine Dengan Representasi TF-IDF

Fahmi Fiddin<sup>1</sup>, Taufik Hidayat<sup>2</sup>, Djamaludin<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Islam Syekh Yusuf, Jln.Syeckh yusuf No.10, Babakan Kec. Tangerang.

Email: <sup>1</sup>fahmifiddin96@gmail.com, <sup>2</sup>thidyat@unis.ac.id, <sup>3</sup>djamaludin@unis.ac.id.

Email Penulis Korespondensi: fahmifiddin96@gmail.com

---

### Article History:

Received Jul 25<sup>th</sup>, 2025

Revised Aug 17<sup>th</sup>, 2025

Accepted Aug 30<sup>th</sup>, 2025

### Abstrak

Penelitian ini mengimplementasikan analisis sentimen pada ulasan produk sayuran di Tokopedia menggunakan metode TF-IDF untuk pembobotan fitur dan algoritma Support Vector Machine (SVM) untuk klasifikasi. Sebanyak 1.048 data ulasan dikumpulkan melalui proses crawling dengan Python dan Selenium. Data ulasan diproses melalui tahapan preprocessing yang mencakup cleaning, case folding, tokenizing, stopword removal, dan stemming. Pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based. Model SVM dilatih dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian, menghasilkan akurasi 98%, F1-Score 98% untuk sentimen positif, dan 96% untuk sentimen negatif. Visualisasi menggunakan Word Cloud dan diagram pie menunjukkan bahwa 75,1% ulasan bersentimen positif dan 24,9% bersentimen negatif. Hasil ini menunjukkan bahwa mayoritas pelanggan memberikan tanggapan positif terhadap produk yang dibeli.

**Kata Kunci** : Analisis Sentimen, Support Vector Machine (SVM), TF-IDF, Ulasan Produk, Tokopedia,

### Abstract

*This study implements sentiment analysis on vegetable product reviews from Tokopedia using Term Frequency-Inverse Document Frequency (TF-IDF) for feature weighting and the Support Vector Machine (SVM) algorithm for classification. A total of 1,048 review data were collected through a crawling process using Python and Selenium. The review data underwent preprocessing stages, including cleaning, case folding, tokenizing, stopword removal, and stemming. Sentiment labeling was performed using a lexicon-based approach. The SVM model was trained with an 80% training and 20% testing data split, achieving an accuracy of 98%, an F1-Score of 98% for positive sentiment, and 96% for negative sentiment. Visualization through Word Cloud and pie charts showed that 75.1% of the reviews were positive, while 24.9% were negative. These results indicate that the majority of customers provided positive feedback on the purchased products.*

**Keyword** : Sentiment Analysis, Support Vector Machine (SVM), TF-IDF , Product Reviews, Tokopedia.

## 1. PENDAHULUAN

Perkembangan teknologi digital dalam beberapa tahun terakhir telah mengubah berbagai aspek kehidupan, termasuk dalam berbelanja kebutuhan pokok seperti sayuran[1]. Seiring dengan kemajuan teknologi, platform *e-commerce* seperti Tokopedia semakin populer di kalangan pelanggan yang ingin membeli barang secara *online*, terutama sayuran segar. Tokopedia merupakan salah satu platform *e-commerce* terbesar di Indonesia yang memberikan kemudahan bagi masyarakat untuk membeli dan menjual barang serta jasa secara daring[2].

Pada tahun 2021, Tokopedia berhasil menduduki posisi pertama dalam jumlah kunjungan dengan perolehan skor sebesar 4,53%[3]. Pada tahun 2023 tokopedia mencatat jumlah kunjungan tertinggi pada kuartal pertama dengan 117,03 juta pengguna aktif. Selain itu, menurut[4] mencatat bahwa Tokopedia merupakan platform *e-commerce* di Indonesia dengan nilai rata-rata transaksi terbesar. Pembelian sayuran melalui platform ini menawarkan kemudahan dan efisiensi, namun tetap menghadirkan tantangan baru, terutama dalam hal kualitas produk yang diterima. Ulasan konsumen

memainkan peran penting dalam membentuk persepsi calon pembeli terhadap produk sayuran yang ditawarkan oleh penjual di Tokopedia[5] Salah satu toko yang beroperasi di Tokopedia adalah toko prima.

Toko Prima adalah toko yang menyediakan layanan pengiriman bahan makanan segar langsung ke tempat. Mereka menawarkan berbagai produk seperti sayuran, buah-buahan, hingga bumbu dapur dengan harga terjangkau. Toko prima juga hadir di platform *e-commerce* seperti Tokopedia. Toko prima sudah mencapai 81 ribu ulasan di tokopedia.

Ulasan konsumen seringkali mencerminkan kualitas produk yang diterima, termasuk kondisi sayuran seperti kesegaran dan ukuran, serta kualitas layanan seperti kecepatan pengiriman. Ulasan yang baik dapat meningkatkan keyakinan calon pembeli untuk melakukan transaksi, sementara ulasan buruk dapat menyebabkan mereka mengurungkan niat membeli produk tersebut[6] Oleh karena itu, ulasan menjadi sumber informasi yang penting bagi calon pembeli untuk memahami pengalaman konsumen lain terhadap produk yang ingin dibeli. Namun, dengan meningkatnya jumlah ulasan pada platform *e-commerce*, menganalisis ulasan secara manual menjadi semakin tidak praktis dan tidak efisien. Salah satu alasan utama mengapa analisis sentimen penting adalah karena analisis sentimen memungkinkan pemrosesan otomatis ulasan teks untuk menentukan sentimen, positif atau negatif, yang terkandung di dalamnya. Tujuan dari analisis ini adalah untuk memahami pandangan konsumen terhadap produk atau layanan dengan memanfaatkan pemrosesan teks secara otomatis. Dalam konteks penjualan sayuran di toko *prima* yang ada di *e-commerce* Tokopedia, analisis sentimen dapat memberikan informasi berharga terkait tingkat kepuasan konsumen terhadap kualitas produk, kecepatan pengiriman, serta pengalaman berbelanja secara keseluruhan[7].

Analisis sentimen dapat berguna bagi toko prima. Bagi toko prima, analisis ini dapat membantu mereka memahami opini konsumen terhadap produk yang dijual, sehingga dapat dijadikan dasar untuk meningkatkan kualitas produk dan layanan. Sementara itu, toko *prima* dapat memanfaatkan hasil analisis ini untuk merekomendasikan produk unggulan kepada calon pembeli berdasarkan ulasan positif dan mengidentifikasi penjual dengan reputasi yang baik. Selain itu, analisis sentimen mendukung toko *prima* dalam meningkatkan pengalaman belanja pengguna dengan menghadirkan produk serta penjual yang sesuai dengan ekspektasi konsumen [8].

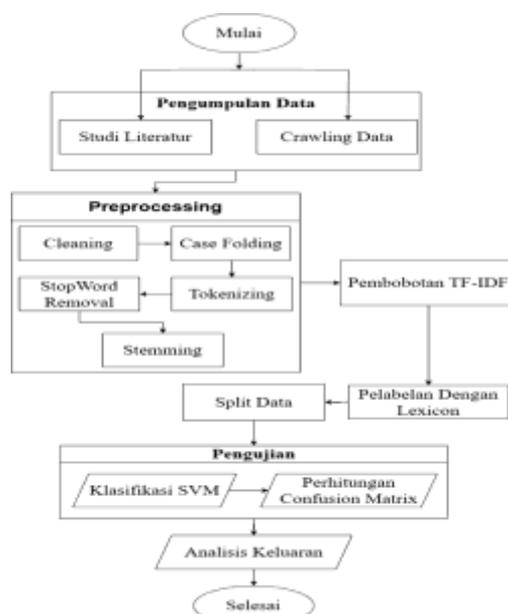
Penelitian sebelumnya yang dilakukan oleh [9] menganalisis sentimen ulasan produk pada *e-commerce*. Menggunakan metode pengembangan sistem, yang mencakup perancangan, implementasi algoritma *naive bayes*, pelatihan, dan pengujian model. Hasilnya, sistem berhasil dikembangkan dengan akurasi 90,91%, yang mengatasi masalah analisis manual dengan cara yang lebih efisien.

Pada penelitian ini akan digunakan analisis sentimen menggunakan pendekatan yang menggabungkan TF-IDF dan *Support Vector Machine (SVM)* diharapkan mampu memberikan model yang akurat dan efisien dalam mengklasifikasi sentimen ulasan produk. penelitian ini bertujuan untuk mengembangkan model analisis sentimen yang akurat dan andal dalam mengklasifikasi sentimen ulasan pembelian sayuran di *e-commerce* Tokopedia.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian yang dirancang sebagai panduan untuk membantu peneliti dalam menjalankan penelitian ini secara sistematis.



Gambar 1. Tahapan Penelitian

## 2.2 Pengumpulan Data

Seperti yang disebutkan di bawah ini, peneliti mengumpulkan berbagai macam data dan informasi yang relevan untuk membantu proses penelitian.

### a. Studi Literatur

Mengumpulkan dan menganalisis buku, makalah, dan jurnal ilmiah yang telah membahas topik sebelumnya di bidang-bidang seperti teori analisis sentimen yang relevan dengan penulisan tesis adalah yang dikenal sebagai tinjauan literatur metode *Support Vector Machine (SVM)*, pengaplikasian bahasa pemrograman Python dalam analisis sentimen [10]. Peneliti juga menggunakan jurnal sebagai referensi pada penulisan skripsi ini sehingga membantu peneliti dalam menyelesaikan penelitian ini.

### b. Crawling Data

Penelitian ini melakukan proses *crawling* data dengan mengumpulkan ulasan produk sayuran dari toko prima. Pengambilan data dilakukan menggunakan *library* Selenium dengan bahasa pemrograman Python dalam rentang waktu 16 Oktober 2024 hingga 2 Desember 2024. Selama proses tersebut, ulasan berhasil dikumpulkan dan disimpan dalam format CSV untuk keperluan analisis lebih lanjut.

## 2.1.2 Preprocessing

Untuk memastikan bahwa algoritma pembelajaran mesin dapat menganalisis data berbasis teks secara efisien, pra pemrosesan teks adalah langkah penting dalam analisis data berbasis teks. Pembersihan, pengorganisasian, dan perapihan data teks mentah merupakan tujuan dari serangkaian langkah prosedur ini. Proses-proses ini bertujuan untuk membersihkan data sehingga model pembelajaran mesin dapat memanfaatkannya dengan lebih baik [11]. Untuk preprocessing teks mencakup beberapa tahapan penting:

### a. Cleaning

Tahap pembersihan (*cleansing*) adalah langkah penting dalam pengolahan teks yang bertujuan untuk menghapus karakter-karakter yang tidak termasuk dalam alfabet. Fokus utama dari teknik ini adalah menghapus informasi yang tidak relevan atau tidak perlu, seperti angka, simbol (\*, \$, dan +), atau bahkan URL yang mungkin disertakan dalam ulasan [12].

### b. Case folding

Pada tahap ini, semua huruf dalam teks akan diubah menjadi huruf kecil. Tujuan dari pendekatan ini adalah untuk memastikan konsistensi saat memproses konten tekstual. Agar perbedaan antara huruf besar dan kecil tidak mempengaruhi analisis, terutama dalam kasus di mana algoritma menganggap keduanya sebagai entitas yang berbeda. Ini adalah langkah penting untuk merampingkan data dan meningkatkan ketepatan analisis teks [13].

### c. Tokenizing

Dengan membagi teks menjadi bagian-bagian yang lebih kecil yang disebut token, proses yang disebut tokenisasi dapat diterapkan. Kata, frasa, atau elemen tekstual lainnya dapat berfungsi sebagai token [14]. Tahap ini sangat krusial dalam analisis teks, termasuk analisis sentimen, karena token berfungsi sebagai unit dasar yang memungkinkan algoritma untuk memahami, menganalisis, dan mengekstrak informasi penting dari teks. Dalam konteks analisis sentimen, token mempermudah proses ekstraksi fitur yang membantu dalam mengidentifikasi pola, emosi, atau opini yang terkandung dalam teks, sehingga meningkatkan akurasi model dalam menghasilkan kesimpulan [15].

### d. Stopword Removal

Salah satu langkah terpenting dalam memproses teks adalah menghapus *stopwords*. Tujuan dari tahap ini adalah untuk menghilangkan istilah yang sering digunakan tetapi tidak benar-benar menambahkan apa pun pada analisis. Kata-kata ini, yang dikenal sebagai *stopwords*, meliputi istilah umum seperti “di”, “ke”, “dan”, dan kata-kata lain yang biasanya tidak menambah informasi dalam konteks analisis [16].

### e. Stemming

*Stemming* adalah proses inti dalam Pemrosesan Bahasa Alami (NLP) yang bertujuan untuk mengubah kata-kata berafiks kembali ke bentuk dasar atau kata akar aslinya. Tujuannya adalah menyamakan berbagai variasi kata yang sebenarnya memiliki makna leksikal yang sama. Proses ini dilakukan dengan menghapus imbuhan yang melekat pada kata. Imbuhan ini bisa berupa awalan (prefiks), sisipan (infiks), atau akhiran (sufiks) [17].

## 2.1.3 Term Frequency – Invers Document Frequency (TF-IDF)

Metode pemrosesan bahasa alami yang disebut TF-IDF (*Term Frequency-Inverse text Frequency*) digunakan untuk menentukan tingkat kepentingan atau bobot setiap kata dalam teks, terutama setelah karakteristik diambil [18]. Untuk menemukan kata-kata yang relevan atau memiliki arti khusus, metode ini membandingkan tingkat kepentingan suatu istilah dalam satu dokumen dengan keseluruhan kumpulan dokumen. Ada dua langkah utama dalam proses TF-IDF. Yang pertama adalah menentukan Frekuensi kata (TF), atau seberapa sering sebuah kata muncul dalam sebuah teks.

Menurut[19] nilai TF yang lebih tinggi mengindikasikan bahwa istilah tersebut lebih sering muncul dalam dokumen yang diberikan.

Langkah kedua adalah *Inverse Document Frequency (IDF)*, yang mengukur seberapa sering sebuah istilah muncul di semua teks[20]. Nilai IDF suatu istilah turun ketika frekuensi dokumennya meningkat, menunjukkan bahwa kata tersebut menjadi kurang khusus atau luas cakupannya. Kata dengan nilai IDF tinggi adalah kata yang jarang muncul di teks lain yang menandakan bahwa kata tersebut memiliki makna khusus dalam konteks dokumen tersebut. Dengan menghitung nilai TF dan IDF secara bersamaan, TF-IDF mengubah kata-kata menjadi representasi numerik yang menggambarkan tingkat kepentingannya. Metode ini banyak digunakan dalam klasifikasi teks, pencarian informasi, dan ekstraksi fitur teks karena kemampuannya dalam membedakan kata-kata yang bermakna dari kata-kata yang lebih umum[21]. Menurut[22] Persamaan untuk *Inverse Document Frequency* adalah sebagai berikut :

$$IDF = \log(N/DFi) + 1 \quad (1)$$

Keterangan:

$N$  = jumlah keseluruhan data

$DFi$  = jumlah dokumen  $d$  yang mengandung kata  $wi$

Sedangkan persamaan untuk *Term Frequency – Inverse Document Frequency* adalah sebagai berikut:

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

Keterangan:

$d$  = dokumen

$IDF(wi)$  = jumlah dokumen  $d$  yang mengandung kata  $wi$

$TF(wi, d)$  = jumlah kemunculan kata  $wi$  di dalam dokumen  $d$

$wi$  = kata ke- $i$

#### 2.1.4 Labeing dengan lexicon

Setelah data ulasan telah diproses dan diberi bobot, kita dapat mengklasifikasikan sentimennya sebagai positif atau negatif. Pendekatan yang paling umum untuk pelabelan kelas termasuk menggunakan kamus atau leksikon untuk mengkategorikan teks, atau mengevaluasi berdasarkan nilai atau peringkat yang diberikan. Kami memutuskan untuk menggunakan kamus leksikal untuk menentukan apakah sebuah ulasan itu baik atau negatif untuk penyelidikan ini.

#### 2.1.5 Split Data

proses ini adalah membagi dataset menjadi dua, membuat dataset yang terpisah untuk pelatihan dan pengujian. Sejalan dengan teknik yang telah ditetapkan sebelumnya, komponen-komponen ini akan digunakan dalam proses klasifikasi.

#### 2.1.6 Support Vector Machine

*Support Vector Machine (SVM)* adalah Menggunakan fungsi linier untuk prediksi atau klasifikasi, SVM mempartisi data dalam ruang fitur berdimensi tinggi untuk pembelajaran mesin. Menemukan *hyperplane* atau batas pemisah yang dapat memisahkan data secara paling efisien menjadi dua kelompok dengan margin terbesar adalah ide inti dari SVM[23]. Proses pelatihan SVM didasarkan pada teori optimasi, yang bertujuan memastikan bahwa jarak antara data dari kelas-kelas yang dipisahkan oleh *hyperplane* tersebut sejauh mungkin[24]. Hal ini bertujuan untuk meminimalkan kemungkinan kesalahan klasifikasi. Bentuk umum *hyperplane* pemisah dapat ditulis seperti pada Persamaan sebagai berikut:

$$W \cdot X + b = 0 \quad (3)$$

keterangan:

$W$  = menyatakan bobot vektor

$X$  = menyatakan data

$b$  = menyatakan nilai bias

Dengan fleksibilitasnya, SVM sangat efektif dalam menangani berbagai jenis data dan permasalahan klasifikasi yang kompleks, seperti pengenalan pola, pengolahan citra, dan analisis teks [25]. Dalam penelitian ini, SVM digunakan untuk mengklasifikasikan sentimen dari ulasan produk sayur di Tokopedia.

### 2.1.7 Confusion Matrix

Salah satu cara untuk membandingkan keakuratan prediksi sistem dengan ground truth dalam hal kategorisasi adalah dengan menggunakan matriks kebingungan. Tabel ini meringkas kinerja model klasifikasi dengan menunjukkan % data uji yang berhasil dikategorikan dan % yang salah diklasifikasikan. Jumlah data positif yang akurat diwakili oleh *True Positive* (TP), jumlah data negatif yang akurat oleh *True Negative* (TN), dan jumlah data negatif oleh *False Positive* (FP) dan data positif oleh *False Negative* (FN). Biasanya, ada empat bagian utama pada tabel ini. Dengan struktur ini, *confusion matrix* sangat berguna untuk mengevaluasi kinerja model klasifikasi dan untuk menemukan pola kesalahan yang bisa diperbaiki[5].

Tabel 1. *Confusion Matrix*

Actual	Prediction	
	Positive	Negative
Positive	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Negative	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Berikut adalah keterangannya:

- TN (*True Negative*) = data kelas aktual bernilai negatif dan model prediksi kelas bernilai negatif.
- FP (*False Positive*) = data kelas aktual bernilai positif dan model prediksi kelas bernilai negatif.
- TP (*True Positive*) = data kelas aktual bernilai positif dan model prediksi bernilai positif.
- FN (*False Negative*) = data kelas aktual bernilai negatif dan nilai prediksi positif.

Confusion matrix menghasilkan 5 metrik perhitungan, yaitu *recall*, *precision*, *accuracy*, dan *f-score*. Menurut [26] pengukuran ini memungkinkan untuk menemukan banyak metrik yang dapat digunakan untuk mengevaluasi kualitas model.

*Recall* mengukur kemampuan model untuk mendeteksi contoh positif. Jika dibandingkan dengan jumlah kasus positif yang diharapkan, rasio perkiraan positif aktual terhadap total kasus positif memberikan nilai ini. Dengan menggunakan rumus sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

*Precision* digunakan untuk mengukur proporsi instance relevan di antara semua instance yang diprediksi positif. Persentase prediksi positif yang benar relatif terhadap jumlah total prediksi positif digunakan untuk menghitungnya. Dengan menggunakan rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Untuk mengukur akurasi, ambil jumlah prediksi yang benar dan membaginya dengan jumlah total perkiraan. Metrik ini menggambarkan tingkat keberhasilan model dalam memprediksi suatu kondisi. Dengan rumus sebagai berikut:

$$Accuracy = \frac{TP + TN}{N} \tag{6}$$

Dalam hal metrik, *F-Score* merupakan perpaduan yang baik antara *recall* dan akurasi. Angka ini dihitung dengan mengambil rata-rata harmonik dari *recall* dan akurasi. Dengan menggunakan rumus sebagai berikut:

$$F - Score = \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pada penelitian ini data-data yang digunakan berasal dari data salah satu toko online yang beroperasi di tokopedia. Total data yang diambil dari periode yang telah ditentukan sebanyak 1048 data. Dalam crawling data ini, menggunakan pemrograman bahasa python dengan libray selenium untuk mengambil data tersebut setelah itu disimpan dalam file format csv. Tabel 2 hasil crawling data yang telah dilakukan.

Tabel 2. Contoh hasil *crawling* data

NO	Ulasan
1.	Baru Pertama kali, Belanja di sini, dan saya puas dengan Pelayanan nya, sayuran nya segar hargapun tidak mahal berikut bapa Grab nya Pun ramah sangat sukaa pokonya mak.
2.	langganan beli buah dan sayur disini produknya fresh pengirimannya cepat dan aman
3.	pesanan jelek sayurannya busuk
4.	sayuran yang layu dan busuk
1044	seller gercep dan cukup tanggap dan yg pasti amanah, mantap
1045	mantap bagus tambah lagi dong jenis sayurannya
1046	sayuran yang busuk dan sangat layu
1047	semua barang lengkap dan baik pengalaman pertama belanja di sini dan puas thank you seller
1048	senang belanja disini harga lebih murah drpd di pasar modern dekat sini sayurannya juga bagus gak nyesel belinya walaupun online terima kasih seller

### 3.2 Preprocessing

Berdasarkan hasil *preprocessing* yang ditampilkan pada Tabel 3, dapat disimpulkan bahwa teks ulasan telah melalui sejumlah tahapan pembersihan dan transformasi untuk mempersiapkannya dalam proses analisis lanjutan. Langkah awal dilakukan dengan membersihkan teks asli guna menghapus karakter yang tidak diperlukan maupun kesalahan penulisan. Setelah itu, seluruh teks dikonversi menjadi huruf kecil guna menjaga konsistensi. Selanjutnya, teks dipecah menjadi kata-kata terpisah agar lebih mudah dianalisis. Pada tahap berikutnya, dilakukan penghapusan kata-kata umum yang tidak memberikan kontribusi berarti terhadap analisis, seperti kata penghubung atau kata yang sering muncul namun kurang bermakna. Terakhir, proses stemming diterapkan untuk mengembalikan kata ke bentuk dasarnya, sehingga berbagai variasi kata yang memiliki makna serupa dapat disederhanakan.

Tabel 3. Contoh hasil *Preprocessing*

Tahapan	Hasil
Ulasan Awal	Baru Pertama kali, Belanja di sini, dan saya puas dengan Pelayanan nya, sayuran nya segar hargapun tidak mahal berikut bapa Grab nya Pun ramah sangat sukaa pokonya mak.
<i>Cleaning</i>	Baru Pertama kali Belanja di sini dan saya puas dengan Pelayanan nya sayuran nya segar hargapun tidak mahal berikut bapa Grab nya Pun ramah sangat sukaa pokonya mak
<i>Case Folding</i>	baru pertama kali belanja di sini dan saya puas dengan pelayanan nya sayuran nya segar hargapun tidak mahal berikut bapa grab nya pun ramah sangat sukaa pokonya mak
<i>Tokenizing</i>	['baru', 'pertama', 'kali', 'belanja', 'di', 'sini', 'dan', 'saya', 'puas', 'dengan', 'pelayanan', 'nya', 'sayuran', 'nya', 'segar', 'hargapun', 'tidak', 'mahal', 'berikut', 'bapa', 'grab', 'nya', 'pun', 'ramah', 'sangat', 'sukaa', 'pokonya', 'mak']
<i>Stopword Removal</i>	['kali', 'belanja', 'puas', 'pelayanan', 'nya', 'sayuran', 'nya', 'segar', 'hargapun', 'mahal', 'bapa', 'grab', 'nya', 'ramah', 'sukaa', 'pokonya', 'mak']
<i>Stemming</i>	['kali', 'belanja', 'puas', 'layan', 'nya', 'sayur', 'nya', 'segar', 'harga', 'mahal', 'bapa', 'grab', 'nya', 'ramah', 'sukaa', 'poko', 'mak']

### 3.3 Pembobotan TF-IDF

Pada tahap ini, dilakukan proses ekstraksi fitur menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency) untuk mengubah data ulasan pelanggan menjadi representasi numerik. Proses ini penting karena algoritma klasifikasi memerlukan data dalam bentuk angka, bukan teks mentah. Dalam penelitian ini, fitur yang diambil adalah 10 kata kunci dengan rata-rata nilai bobot TF-IDF tertinggi dalam korpus ulasan. Kata-kata tersebut adalah segar, sayur, busuk, layu, bagus, langgan, terima, kasih, kirim dan cepat. Setiap kata diberikan bobot berdasarkan tingkat kepentingannya dalam suatu ulasan dibandingkan dengan ulasan lain. Nilai TF-IDF yang tinggi menunjukkan bahwa kata tersebut sering muncul pada ulasan tertentu, namun jarang ditemukan pada ulasan lain, sehingga memiliki nilai informasi yang tinggi. Sebaliknya, jika nilai TF-IDF sebuah kata adalah 0, maka kata tersebut tidak muncul dalam ulasan tersebut. Tabel 4 berikut menunjukkan hasil transformasi TF-IDF pada beberapa data ulasan. Dapat dilihat bahwa distribusi nilai TF-IDF berbeda pada setiap baris (ulasan), yang mencerminkan pentingnya kata-kata tertentu dalam konteks ulasan yang berbeda.

Tabel 4. hasil *TF-IDF*

No	Hasil	Weight
1	segar	0.104120
2	sayur	0.098197
3	busuk	0.086716
4	layu	0.080086
5	bagus	0.065907
6	langgan	0.039920
7	terima	0.039297
8	kasih	0.038987
9	kirim	0.035235
10	cepat	0.033329

### 3.4 Labeling Data

Setelah melalui tahapan preprocessing dan penerapan TF-IDF pada data ulasan pelanggan, langkah berikutnya adalah melakukan pelabelan sentimen dengan menggunakan pendekatan berbasis leksikon (lexicon-based). Proses ini dilakukan dengan mencocokkan setiap token hasil stemming dari ulasan dengan daftar kata yang terdapat dalam lexicon yang telah disesuaikan dan diperluas untuk konteks analisis ulasan produk sayuran. Hasil dari proses pelabelan tersebut ditampilkan pada Tabel 5.

Tabel 5. hasil *Labeling Data*

No	Ulasan	Label	Score
1	baru pertama kali belanja di sini dan saya puas dengan pelayanannya sayuran nya segar hargapun tidak mahal berikut bapa grab nya pun ramah sangat sukaa pokonya mak	<i>Positive</i>	9
2	pesan jam pagi sampai jam di tangerang kota packing oke sangat sayuran segar semua besar dan bagus recommended items seller semoga varian sayuran lainnya	<i>Positive</i>	15
3	langganan beli buah dan sayur disini produknya baik pengirimannya cepat dan aman	<i>Positive</i>	13
4	pesanan jelek sayurannya busuk	<i>Negative</i>	-22
5	sayuran yang layu dan busuk	<i>Negative</i>	-23

### 3.5 Klasifikasi dengan Support Vector Machine

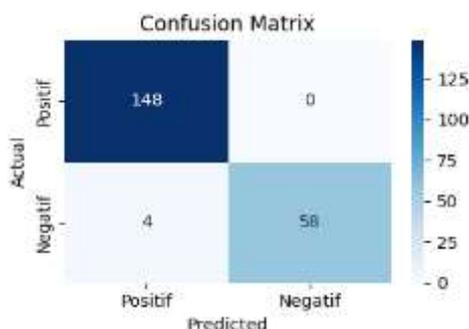
Pada tahap ini, data yang telah diproses melalui tahapan preprocessing, pembobotan dengan TF-IDF, serta pelabelan sentimen, kemudian diklasifikasikan menggunakan algoritma Support Vector Machine (SVM). Pemilihan algoritma SVM didasarkan pada kemampuannya yang unggul dalam mengolah data teks skala besar serta membedakan kelas dengan margin pemisahan yang optimal. Proses pelatihan dilakukan dengan membagi dataset menjadi 80% untuk data pelatihan dan 20% untuk data pengujian. Model yang telah dilatih kemudian dievaluasi menggunakan data uji, guna menilai performanya dalam mengklasifikasikan ulasan ke dalam dua kategori sentimen, yaitu positif dan negatif. Adapun hasil klasifikasi diperoleh sebagai berikut:

Tabel 5. hasil Klasifikasi dengan *Support Vector Machine*

Label	Precision	Recall	F-score	Support
Positif	97.37%	100.00%	98.67%	148
Negatif	100.00%	93.55%	96.67%	62
Akurasi		98%		

### 3.6 Evaluasi Model

Evaluasi terhadap model dilakukan dengan memanfaatkan confusion matrix untuk mengukur kinerja klasifikasi sentimen yang dihasilkan oleh algoritma *Support Vector Machine* (SVM). *Confusion matrix* memberikan gambaran mengenai jumlah prediksi yang tepat maupun yang keliru, berdasarkan kategori sentimen positif dan negatif. Dari matriks tersebut, diperoleh empat komponen utama, yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Visualisasi hasil *confusion matrix* ditampilkan pada Gambar 2.



Gambar 2. *Confusion Matrix*

Berdasarkan pada Gambar 2, untuk kelas positif, model mampu mengklasifikasikan sebanyak 148 data dengan benar. Tidak ditemukan adanya data positif yang salah diklasifikasikan sebagai negatif (*false negative* = 0). Sementara itu, pada kelas negatif, sebanyak 58 data berhasil diklasifikasikan secara tepat, meskipun masih terdapat 4 data negatif yang keliru diprediksi sebagai positif (*false positive*).

### 3.7 Word cloud positive

Word Cloud digunakan untuk memvisualisasikan kata-kata yang paling sering muncul dalam ulasan dengan sentimen positif. Gambar 2 menampilkan hasil Word Cloud untuk kategori ini.



Gambar 3. *Word cloud positive*

Berdasarkan gambar, dapat dilihat bahwa kata-kata seperti "bagus", "segar", "seller", "langganan", "sayur", dan "kirim" muncul dengan ukuran yang paling besar dan menonjol. Ini menunjukkan bahwa kata-kata tersebut paling sering dan dominan digunakan oleh pelanggan dalam memberikan ulasan positif. Secara spesifik, penonjolan kata "bagus" dan "segar" mengindikasikan kepuasan pelanggan terhadap kualitas produk. Kata "seller" dan "langganan" menunjukkan apresiasi terhadap pelayanan penjual dan keinginan untuk menjadi pelanggan tetap. Kemunculan kata "kirim" dengan ukuran besar menyoroti pentingnya kecepatan pengiriman bagi pelanggan.

Selain itu, kata-kata seperti "cepat", "packing", dan "bersih" juga sering muncul dengan ukuran signifikan, menggarisbawahi bahwa pelanggan sangat menghargai kecepatan pengiriman, kerapian pengemasan produk, serta kondisi produk yang bersih. Meskipun tidak sebesar kata-kata yang disebutkan sebelumnya, kemunculan kata-kata seperti



Berdasarkan hasil penelitian mengenai analisis sentimen pada ulasan pelanggan produk sayuran di platform Tokopedia, yang memanfaatkan metode TF-IDF dan algoritma Support Vector Machine (SVM), diperoleh kesimpulan bahwa pengumpulan data dilakukan melalui teknik crawling dengan bahasa pemrograman Python serta library Selenium, sehingga terkumpul sebanyak 1.048 data ulasan dari salah satu toko di Tokopedia. Tahapan preprocessing berhasil mengubah teks ulasan mentah menjadi format yang siap dianalisis, melalui serangkaian proses seperti pembersihan (cleaning), konversi huruf (case folding), tokenisasi (tokenizing), penghapusan kata umum (stopword removal), serta stemming. Proses ekstraksi fitur berbasis TF-IDF mampu mengidentifikasi kata-kata kunci dengan bobot tertinggi, seperti "segar", "sayur", "busuk", dan "layu", yang merepresentasikan isi ulasan secara umum.

Selanjutnya, pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based, yang secara otomatis mengklasifikasikan ulasan ke dalam kategori sentimen positif atau negatif berdasarkan daftar kata yang telah disesuaikan untuk konteks produk sayuran. Hasil klasifikasi menggunakan algoritma SVM menunjukkan performa yang sangat baik, dengan tingkat akurasi mencapai 98%. Selain itu, F1-Score untuk sentimen positif mencapai 98%, sementara untuk sentimen negatif sebesar 96%, berdasarkan evaluasi melalui confusion matrix. Visualisasi data dalam bentuk word cloud dan diagram pie turut memperkuat gambaran distribusi sentimen serta kata-kata yang dominan pada masing-masing kategori. Dari hasil analisis tersebut diketahui bahwa 75,1% ulasan memiliki sentimen positif, sedangkan 24,9% lainnya bersentimen negatif.

## UCAPAN TERIMA KASIH

Penulis mengucapkan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan, bantuan, serta kontribusi dalam berbagai bentuk, sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih juga ditujukan kepada dosen pembimbing, rekan-rekan sejawat, serta semua pihak yang secara langsung maupun tidak langsung telah membantu dalam proses penyusunan dan pelaksanaan penelitian ini.

## DAFTAR PUSTAKA

- [1] J Rajeev, N. suresh, and T. Varalakshmi, "Sentiment Analysis About Customer Feedback," *Int. Res. J. Adv. Eng. Manag.*, vol. 2, no. 05, pp. 1683–1686, 2024, doi: 10.47392/irjaem.2024.0242.
- [2] K. S. Kyaw, P. Tepsongkroh, and C. Thongkamkaew, "Business Intelligent Framework Using Sentiment Analysis for Smart Digital Marketing in the E-Commerce Era Department of Tourism Industry , Didyardin International College , Department of Management , Hatyai Business School , Hatyai University , Correspon," vol. 16, no. 3, pp. 1–23, 2022, doi: <https://doi.org/10.48048/asi.2023.252965>.
- [3] T. Hidayat and D. S. Canta, "Analisis Kepuasan Pengguna Terhadap Penerapan Aplikasi Tokopedia dengan Menggunakan Metode TAM," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 472, 2022, doi: <http://dx.doi.org/10.30865/jurikom.v9i2.4088>.
- [4] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1329, Oct. 2022, doi: <http://dx.doi.org/10.30865/jurikom.v9i5.4774>.
- [5] M. E. Purbaya, D. P. Rakhmadani, Maliana Puspa Arum, and Luthfi Zian Nasifah, "Implementation of n-gram Methodology to Analyze Sentiment Reviews for Indonesian Chips Purchases in Shopee E-Marketplace," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 609–617, 2023, doi: 10.29207/resti.v7i3.4726.
- [6] A. H. Ruger, M. Suyanto, and M. P. Kurniawan, "Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes," *J. Inf. Technol.*, vol. 1, no. 2, pp. 26–29, Sep. 2021, doi: 10.46229/jifotech.v1i2.282.
- [7] K. Hantoro, D. Handayani, and S. Setiawati, "A Implementation of Text Mining In Sentiment Analysis of Shopee Indonesia Using SVM," *Bull. Inf. Technol.*, vol. 3, no. 2, pp. 115–120, Jun. 2022, doi: 10.47065/bit.v3i2.282.
- [8] Z. Fu, Y. C. Hsu, C. S. Chan, C. M. Lau, J. Liu, and P. S. F. Yip, "Efficacy of ChatGPT in Cantonese Sentiment Analysis: Comparative Study," *J. Med. Internet Res.*, vol. 26, no. 1, p. e51069, Jan. 2024, doi: <https://doi.org/10.2196/51069>.
- [9] A. Muzaki, V. Febriana, and W. N. Cholifah, "Analisis Sentimen Pada Ulasan Produk di E-Commerce dengan Metode Naive Bayes," *J. Ris. dan Apl. Mhs. Inform.*, vol. 5, no. 4, pp. 758–765, Oct. 2024, doi: 10.30998/jrami.v5i4.9647.
- [10] M. A. Nadjib Mahfoud, T. Hidayat, Sukrim, Sukisno, and A. H. Nugroho, "Penerapan Algoritma Dijkstra Dan Greedy Untuk Optimasi Rute Angkut Sampah Di Kecamatan Periuk," vol. 24, no. 2, pp. 151–161, 2024, doi: <https://doi.org/10.33557/jurnalmatrik.v26i2.3259>.
- [11] F. Fiddin, M. Y. Syahbarna, and M. Ridwan, "Penggunaan Supervised Learning untuk Prediksi Validitas Ulasan Negatif Aplikasi Tokopedia Berdasarkan Pengalaman Pengguna Ahli," *SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, pp. 409–417, 2024, doi: <https://doi.org/10.53513/jis.v23i2.10030>.
- [12] R. Ridho and H. Hendra, "Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree," *JUST IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 11, no. 3, pp. 69–75, 2022, doi: <https://doi.org/10.24853/justit.11.3.%25p>.

- [13] B. Blidex and J. S. Wibowo, "ANALISIS SENTIMEN KLASIFIKASI TWEET VAKSIN COVID 19 DENGAN NAÏVE BAYES," *J. MAHAJANA Inf.*, vol. 6, no. 2, pp. 103–110, Dec. 2021, doi: 10.51544/jurnalmi.v6i2.2449.
- [14] R. Ramadhani, R. Ramadhanu, A. Fajri, A. Abdillah, and M. Ridwan, "Studi Komparatif Multinomial Naïve Bayes, Decision Tree, dan K-Nearest Neighbor dalam Klasifikasi Validasi Ulasan Clash of Clans oleh Pengguna Ahli," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 4, pp. 653–660, 2024, doi: 10.26418/justin.v12i4.81638.
- [15] A. K. Santoso, "Analisis Sentimen Twitter Bahasa Indonesia Menggunakan Pendekatan Machine Learning," *J. Inform. Kaputama*, vol. 6, no. 2, pp. 129–136, 2022, doi: 10.59697/jik.v6i2.111.
- [16] F. Hasibuan, W. Priatna, and T. S. Lestari, "Analisis Sentimen Terhadap Kementerian Perdagangan Pada Sosial Media Twitter Menggunakan Metode Naïve Bayes," *Techno.Com*, vol. 21, no. 4, pp. 741–752, 2022, doi: 10.33633/tc.v21i4.6565.
- [17] Muhammad Rizal, M. Martanto, and U. Hayati, "Analisis Sentimen Pengguna Twitter Terkait Film One Piece Menggunakan Metode Naive Bayes," *J. Sist. Inf. Kaputama*, vol. 8, no. 1, pp. 38–47, 2024, doi: 10.59697/jsik.v8i1.522.
- [18] A. Yodi Prayoga, A. Id Hadiana, and F. Rakhmat Umbara, "Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf," *J. Syntax Admiration*, vol. 2, no. 10, pp. 1808–1823, 2021, doi: 10.46799/jsa.v2i10.327.
- [19] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
- [20] Regina, T. H. Saragih, and D. Kartini, "Analisis Sentimen Brand Ambassador Bts Terhadap Tokopedia Menggunakan Klasifikasi Bayesian Network Dengan Ekstraksi Fitur Tf-Idf," *J. Inform. Polinema*, vol. 9, no. 4, pp. 383–390, 2023, doi: 10.33795/jip.v9i4.1333.
- [21] A. Aziz and Fauziah, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, p. 115, 2022, doi: <http://dx.doi.org/10.30645/j-sakti.v6i1.430>.
- [22] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer," *DIKE J. Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, 2023, doi: 10.69688/dike.v1i2.35.
- [23] A. M. Yolanda and R. T. Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *J. Stat. Its Appl. Teach. Res.*, vol. 6, no. 2, pp. 76–83, 2024, doi: 10.35580/variansium258.
- [24] R. Pebrianto, T. Rivanie, R. Nurfalih, W. Gata, and M. F. Julianto, "Adopsi Algorithm Support Vector Machine untuk Analisis Sentimen Larangan Mudik Lebaran 2020 pada Twitter," *J. Tek. Komput. AMIK BSI*, vol. 6, no. 2, pp. 193–199, 2020, doi: 10.31294/jtk.v6i2.8127.
- [25] Sopiatal Ulum, R. F. Alifa, P. Rizkika, and C. Rozikin, "Perbandingan Performa Algoritma KNN dan SVM dalam Klasifikasi Kelayakan Air Minum," *Gener. J.*, vol. 7, no. 2, pp. 141–146, 2023, doi: 10.29407/gj.v7i2.20270.
- [26] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. December 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.