JEMSI: Jurnal Ekonomi Manajemen Sistem Informasi
E-ISSN: 2686-5238
P-ISSN: 2686-4916
https://dinastirev.org/JEMSI dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/jemsi.v7i2
https://creativecommons.org/licenses/by/4.0/
Received: 07 November 2025 | **Revised:** 13 November 2025 | **Accepted:** 23 November 2025

Analisis Sentimen Pelanggan terhadap Penilaian Produk pada Tokopedia *Nyemil.Saji* menggunakan Metode *Support Vector Machine*

Putri Demetria[1], Ari Wedhasmara[2]
[1]Universitas Sriwijaya, Sumatera Selatan, Indonesia, pidipei@gmail.com
[2]Universitas Sriwijaya, Sumatera Selatan, Indonesia, a_wedhasmara@unsri.ac.id

Corresponding Author: a_wedhasmara@unsri.ac.id[1]

Abstract: *This study aims to analyze customer sentiment toward product reviews on the Nyemil.Saji store on the Tokopedia platform by applying the Support Vector Machine (SVM) method. The background of this study is based on the ambiguity between the content of text reviews and star ratings, which can make it difficult to assess product quality objectively. Customer review data was collected through web scraping techniques and processed using text preprocessing stages (case folding, cleansing, tokenizing, stopword removal, stemming, and normalization), followed by Term Frequency–Inverse Document Frequency (TF-IDF) feature representation. A total of 2,826 customer reviews were analyzed in this study to ensure the accuracy of the model. The SVM model was used as the main classification algorithm to determine positive, negative, and neutral sentiment categories. The results showed that the SVM model produced an accuracy rate of 96% with the highest precision and recall values for neutral sentiment, while the highest F1-score was obtained for positive sentiment. These findings indicate that Nyemil.Saji customers tend to give neutral to positive reviews of the products they purchase, which indicates a good level of customer satisfaction with the taste and delivery service.*

Keywords: *Sentiment Analysis, Support Vector Machine, TF-IDF, Tokopedia, E-Commerce.*

Abstrak: Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap ulasan produk pada toko *Nyemil.Saji* di platform Tokopedia dengan menerapkan metode *Support Vector Machine* (SVM). Latar belakang penelitian ini didasarkan pada permasalahan ambiguitas antara isi ulasan teks dan *rating* bintang yang dapat menimbulkan kesulitan dalam menilai kualitas produk secara objektif. Data ulasan pelanggan dikumpulkan melalui teknik *web scraping* dan diproses menggunakan tahapan *text preprocessing (case folding, cleansing, tokenizing, stopword removal, stemming,* dan normalisasi), diikuti dengan representasi fitur *Term Frequency–Inverse Document Frequency* (TF-IDF). Sebanyak 2.826 ulasan pelanggan dianalisis dalam penelitian ini untuk memastikan keakuratan model. Model SVM digunakan sebagai algoritma klasifikasi utama untuk menentukan kategori sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa model SVM menghasilkan tingkat akurasi sebesar 96% dengan nilai *precision* dan *recall* tertinggi pada sentimen netral, sedangkan *F1-score*

tertinggi diperoleh pada sentimen positif. Temuan ini menunjukkan bahwa pelanggan *Nyemil.Saji* cenderung memberikan ulasan yang netral hingga positif terhadap produk yang dibeli, yang menandakan tingkat kepuasan pelanggan yang baik terhadap kualitas rasa dan pelayanan pengiriman.

Kata Kunci: Analisis Sentimen, *Support Vector Machine,* TF-IDF, Tokopedia, *E-commerce.*

PENDAHULUAN

Teknologi informasi telah mempermudah pelaku usaha dalam melakukan transaksi daring, mengelola data, dan berinteraksi dengan pelanggan sehingga mendorong pertumbuhan bisnis berbasis digital seperti *e-commerce* dan *marketplace*. Kehadiran platform *e-commerce* memberikan kemudahan, efisiensi waktu, serta variasi produk yang luas sehingga semakin diminati masyarakat (Ahadiyah, 2024; Clara & Beni, 2023). Di dalam ekosistem *e-commerce*, ulasan atau *review* pelanggan menjadi sumber informasi penting bagi calon pembeli dan pelaku usaha karena merepresentasikan pengalaman pengguna serta dapat memengaruhi keputusan pembelian (Siti, Sarjita, & Edi, 2023; Ayuningrum & Isa, 2024).

Namun praktik pemanfaatan ulasan pelanggan tidak selalu sederhana. Sering ditemukan ketidaksesuaian antara konten teks ulasan dan penilaian kuantitatif berupa *rating* bintang misalnya ulasan naratif yang positif namun disertai *rating* rendah yang menimbulkan ambiguitas dalam menilai kualitas produk dan menyulitkan penjual dalam evaluasi performa toko secara objektif (Muzaki, Febriana, & Cholifah, 2024; Kosasih & Alberto, 2021). Masalah ini menjadi sangat krusial bagi pelaku UMKM yang menerima *volume* ulasan besar sehingga analisis manual tidak lagi efisien. Berdasarkan informasi dari pengelola toko *Nyemil.Saji*, *volume* ulasan yang diterima setiap harinya cukup tinggi, sehingga diperlukan pendekatan otomatis untuk mengolah data tersebut secara efisien.

Analisis sentimen, bagian dari *text mining* yang memanfaatkan teknik *Natural Language Processing*, berperan untuk mengklasifikasikan opini dalam teks menjadi kategori sentimen (positif, negatif, netral) sehingga membantu memahami persepsi konsumen secara lebih objektif dan mendukung pengambilan keputusan bisnis (Kamal & Ratnasari, 2020; Hajaroh, Suprapti, & Narasati, 2024). Dalam penelitian ini, *sentimen* dipahami sebagai kecenderungan opini pelanggan yang tercermin dalam ulasan teks yang dapat dikategorikan sebagai positif, negatif, atau netral berdasarkan isi bahasa dan konteks evaluatif penulis ulasan (Kosasih & Alberto, 2021; Kusuma & Cahyono, 2023).

Sejumlah studi terdahulu menunjukkan variasi sumber data, teknik *preprocessing*, dan algoritma klasifikasi yang digunakan pada analisis sentimen ulasan *e-commerce*. Beberapa penelitian menggunakan *Naïve Bayes* dengan TF-IDF menghasilkan akurasi sedang (Permadi, 2020; Hajaroh et al., 2024; Arsadhana et al., 2025), sementara studi komparatif menemukan *Support Vector Machine* (SVM) memberikan performa unggul pada klasifikasi teks di berbagai konteks (Harahap et al., 2021; Wijaya, Indriati, & Muzaki, 2021). Mengingat sifat data teks yang berdimensi tinggi dan kebutuhan akan model yang stabil dan akurat, metode SVM dipilih untuk penelitian ini karena kemampuannya memisahkan kelas dengan *hyperplane* optimal sehingga sering memberikan performa yang stabil pada tugas klasifikasi teks (Idris, Mustofa, & Salihi, 2023; Fitri et al., 2024).

Berdasarkan uraian tersebut, penelitian ini diarahkan untuk mengatasi ambiguitas antara isi ulasan teks dan *rating* bintang pada platform Tokopedia milik *Nyemil.Saji* melalui pendekatan analisis sentimen berbasis *Support Vector Machine.* Rumusan masalah yang diangkat adalah: Bagaimana pandangan pelanggan terhadap penilaian produk di Tokopedia *Nyemil.Saji* apabila dianalisis menggunakan metode *Support Vector Machine*? Tujuan penelitian ini adalah untuk menetapkan atau mengidentifikasi sentimen pelanggan (positif,

negatif, atau netral) terhadap produk yang dijual oleh *Nyemil.Saji* di Tokopedia dengan menggunakan metode SVM.

Penelitian ini diharapkan memberikan manfaat teoritis berupa kontribusi pada pengembangan kajian data mining dan analisis sentimen dalam konteks UMKM di *platform e-commerce* di Indonesia (Ridho et al., 2023; Afandi et al., 2024), serta manfaat praktis bagi pelaku usaha *Nyemil.Saji* dalam memahami persepsi pelanggan secara lebih mendalam untuk keperluan evaluasi produk, peningkatan layanan, dan strategi pemasaran tanpa bergantung pada analisis manual yang memakan waktu dan rawan bias.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *text mining* untuk menganalisis sentimen pelanggan terhadap produk *Nyemil.Saji* di platform *e-commerce* Tokopedia. Analisis dilakukan menggunakan algoritma *Support Vector Machine* (SVM) sebagai model klasifikasi utama dengan representasi fitur berbasis *Term Frequency–Inverse Document Frequency* (TF-IDF).

Data dan Sampel Penelitian

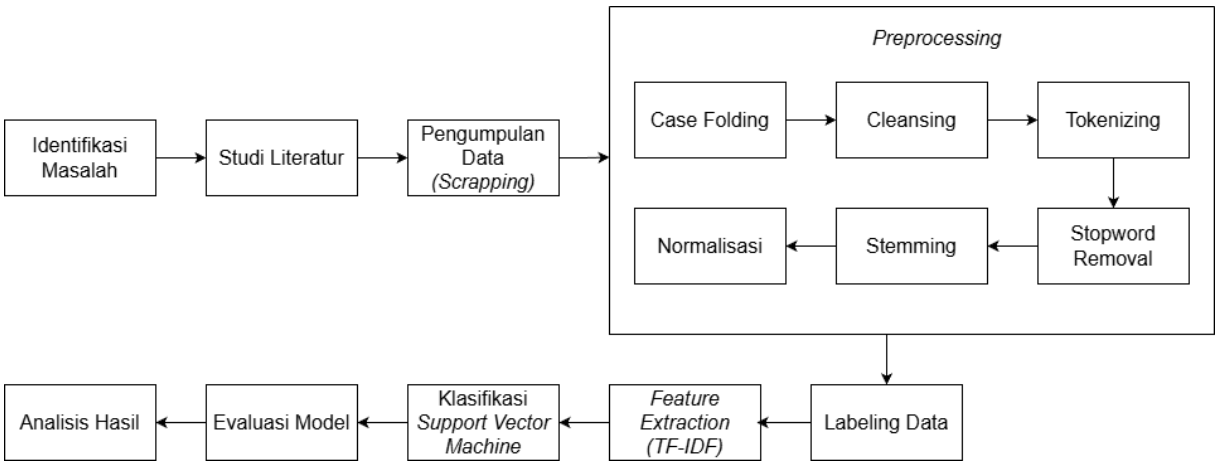
Data yang digunakan dalam penelitian ini merupakan data sekunder berupa teks ulasan pelanggan dan *rating* yang diperoleh dari halaman toko *Nyemil.Saji* di Tokopedia. Pengumpulan data dilakukan menggunakan teknik *web scraping* untuk mengambil ulasan pelanggan secara otomatis melalui bahasa pemrograman *Python* dengan pustaka *BeautifulSoup* dan *Selenium* untuk pemrosesan halaman dinamis, serta *pandas* untuk penyimpanan data dalam format terstruktur (CSV/JSON). Data yang telah diperoleh sebanyak 2.826 kemudian diseleksi menggunakan metode *purposive sampling*, yaitu hanya ulasan yang relevan, memiliki opini jelas, dan mencerminkan persepsi pelanggan yang digunakan dalam analisis (Ulfah & Najiah, 2023; Daruhadi & Sopiati, 2024).

Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada tahun 2025, dengan proses pengumpulan data dilakukan secara daring melalui *platform* Tokopedia, sedangkan proses analisis dan pemodelan dilaksanakan menggunakan perangkat komputer dengan dukungan perangkat lunak *Python* dan pustaka pendukung seperti *scikit-learn* dan *NLTK*.

Prosedur Penelitian

Prosedur penelitian dilakukan melalui beberapa tahap sebagaimana alur pada Gambar 1.



Gambar 1. Alur Penelitian

Tahap pertama adalah identifikasi masalah dan studi literatur, yaitu penelusuran teori dan penelitian terdahulu yang relevan dengan analisis sentimen dan algoritma SVM. Tahap berikutnya adalah pengumpulan data, yang dilakukan menggunakan teknik *web scraping* terhadap ulasan pelanggan di Tokopedia.

Data yang diperoleh kemudian diproses melalui tahap *preprocessing*, yaitu pembersihan dan standarisasi teks agar dapat dianalisis secara komputasional (Harahap, Sihombing, Laia, Saragih, & Dharma, 2021). Tahapan *preprocessing* meliputi :

1. *Case folding* untuk mengubah seluruh karakter menjadi huruf kecil.
2. *Cleansing* untuk menghapus karakter non-alfabet seperti tanda baca, simbol, dan emotikon.
3. *Tokenizing* untuk memecah teks menjadi kata-kata (token).
4. *Stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna sentimen, seperti "yang", "dan", atau "di".
5. *Normalisasi* untuk menyeragamkan kata tidak baku, misalnya "bgt" menjadi "banget".
6. *Stemming* untuk mengembalikan kata ke bentuk dasarnya.

Labelling Data

Proses *labelling* dilakukan secara semi-otomatis berdasarkan kombinasi antara *rating* bintang dan konteks kalimat pada ulasan pelanggan. Penentuan kategori sentimen dilakukan dengan merujuk pada aturan operasional berikut:

Tabel 1. Kriteria *Labelling*

Kategori	Kriteria Utama	Contoh Kata Kunci
Positif	*Rating* 4–5 bintang dan/atau mengandung kata positif kuat	"enak", "puas", "cepat", *"recommended"*
Netral	*Rating* 3 bintang atau teks bersifat informatif tanpa ekspresi emosional kuat	"oke", "sesuai deskripsi", "biasa saja"
Negatif	*Rating* 1–2 bintang dan/atau mengandung kata negatif	"kecewa", "rusak", "lama", "tidak enak"

Jika terjadi ketidaksesuaian antara isi teks dan *rating* (misalnya teks positif dengan *rating* 3), maka keputusan akhir ditentukan berdasarkan konteks kalimat utama yang dievaluasi oleh dua annotator independen untuk menjaga konsistensi hasil pelabelan (Ulfah & Najiah, 2023; Daruhadi & Sopiati, 2024).

Ekstraksi Fitur dan Klasifikasi SVM

Setelah proses pelabelan selesai, dilakukan ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* untuk mengubah teks menjadi representasi numerik (Pratiwi et al., 2021). Hasil ekstraksi ini digunakan sebagai input dalam proses klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM). Model klasifikasi dikembangkan menggunakan pustaka *scikit-learn,* dengan parameter utama sebagai berikut :

```
from sklearn.svm import SVC
model = SVC(kernel='linear', C=1.0, gamma='scale', random_state=42)
```

Gambar 2. Pustaka Parameter Utama

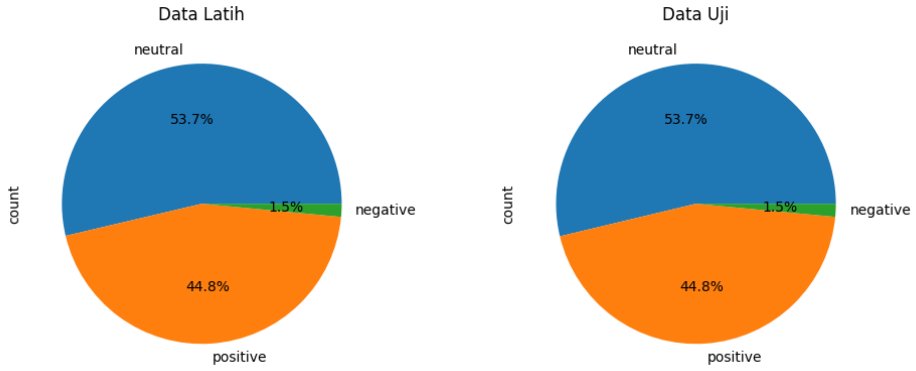
Pemilihan kernel *linear* dilakukan karena data bersifat *linier separable* dan menghasilkan performa terbaik pada tahap uji coba awal. Algoritma ini bekerja dengan mencari *hyperplane* terbaik yang memisahkan kategori sentimen positif, negatif, dan netral (Fitri et al., 2024). Jika data tidak dapat dipisahkan secara linear, SVM akan menggunakan fungsi *kernel* untuk memetakan data ke ruang berdimensi lebih tinggi (Pratiwi et al., 2021).

Tabel 2. Parameter Model SVM

Parameter	Nilai	Keterangan
Kernel	Linear	Pemisah linier antar kelas
C	1.0	Tingkat regularisasi
Gamma	Scale	Skala otomatis untuk fitur *TF-IDF*
Random State	42	Reproduksibilitas hasil

Split Data

Dataset dibagi menjadi dua bagian, yaitu data latih *(training data)* dan data uji *(testing data)* dengan proporsi yang seimbang antara setiap label sentimen. Data latih digunakan untuk melatih model SVM agar dapat mengenali pola sentimen, sedangkan data uji digunakan untuk mengukur performa model terhadap data baru yang belum pernah dilihat sebelumnya Ayuningrum & Isa, 2024). Distribusi label pada data latih dan data uji ditunjukkan pada Gambar 3 berikut.



Gambar 3. Pembagian Data Latih dan Data Uji

Berdasarkan distribusi tersebut, diketahui bahwa pada data latih terdapat 1.604 data, yang terdiri dari 861 label *neutral* (53,7%), 719 label *positive* (44,8%), dan 24 label *negative* (1,5%). Sedangkan pada data uji terdapat 402 data, dengan komposisi 216 label *neutral* (53,7%), 180 label *positive* (44,8%), dan 6 label *negative* (1,5%).

Pembagian data ini dilakukan dengan mempertahankan proporsi seimbang antar kelas *(stratified split)* sehingga distribusi sentimen pada data latih dan data uji tetap konsisten. Dengan demikian, model dapat dilatih dan diuji secara representatif terhadap keseluruhan populasi data.

Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix* untuk menilai performa klasifikasi berdasarkan empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (Hajaroh, Suprapti, & Narasati, 2024).

Tabel 3. *Confusion Matrix*

Kelas Aktual \ Kelas Prediksi	Negatif	Netral	Positif
Negatif	2	4	0
Netral	0	212	4
Positif	0	8	172

Metrik-metrik tersebut digunakan untuk mengukur sejauh mana model SVM mampu memprediksi kategori sentimen pelanggan secara tepat. Berdasarkan hasil *confusion matrix* di atas, dapat diketahui sebaran prediksi untuk masing-masing kelas sentimen. Hasil evaluasi ini kemudian digunakan untuk menganalisis kecenderungan opini pelanggan serta proporsi sentimen dominan terhadap produk *Nyemil.Saji* di Tokopedia.

HASIL DAN PEMBAHASAN

Preprocessing

Proses *preprocessing* dilakukan untuk menyiapkan data teks hasil *web scraping* agar dapat diolah dengan algoritma klasifikasi. Setiap tahapan menghasilkan keluaran data yang lebih bersih dan terstruktur sebagaimana ditunjukkan pada Tabel 4 hingga Tabel 10.

Case Folding

Tahap *case folding* bertujuan untuk menyeragamkan seluruh teks menjadi huruf kecil *(lowercase)* sehingga sistem tidak membedakan huruf besar dan kecil dalam proses analisis.

Tabel 4. Hasil *Case Folding*

No	Sebelum Proses	Sesudah Proses
1	ENAK BANGET!!! Produk DIKIRIM CEPAT dan Recommended 100% 😋	enak banget!!! produk dikirim cepat dan recommended 100% 😋
2	Gurih Sekali, tapi agak mahal 😅	gurih sekali, tapi agak mahal 😅

Tahapan ini menghasilkan teks dengan format huruf kecil yang lebih seragam sehingga memudahkan proses pembersihan karakter non-alfabet pada tahap berikutnya.

Cleansing

Tahap *cleansing* dilakukan untuk menghapus karakter yang tidak diperlukan seperti angka, tanda baca, emotikon, tautan, dan simbol agar teks bersih dari *noise.*

Tabel 5. Hasil *Cleansing*

No	Sebelum Proses	Sesudah Proses
1	enak banget!!! produk dikirim cepat dan recommended 100% 😋	enak banget produk dikirim cepat dan recommended
2	gurih sekali, tapi agak mahal 😅	gurih sekali tapi agak mahal

Hasil pada Tabel 5 menunjukkan bahwa setelah pembersihan, teks menjadi lebih bersih dan siap diolah secara linguistik.

Tokenizing

Tahapan *tokenizing* memecah kalimat menjadi potongan-potongan kata yang disebut *token*, agar setiap kata dapat dianalisis secara individual.

Tabel 6. Hasil *Tokenizing*

No	Sebelum Proses	Sesudah Proses
1	enak banget produk dikirim cepat dan recommended	[“enak”, “banget”, “produk”, “dikirim”, “cepat”, “dan”, “recommended”]
2	gurih sekali tapi agak mahal	[“gurih”, “sekali”, “tapi”, “agak”, “mahal”]

Tokenisasi membantu memecah struktur kalimat menjadi elemen kata yang dapat diberi bobot dalam proses ekstraksi fitur TF-IDF.

Stopword Removal

Tahapan ini menghapus kata-kata umum yang sering muncul namun tidak memiliki makna penting terhadap konteks sentimen, seperti “yang”, “di”, “dan”, “ke”, dan sebagainya.

Tabel 7. Hasil *Stopword Removal*

No	Sebelum Proses	Sesudah Proses

1	[“enak”, “banget”, “produk”, “dikirim”, “cepat”, “dan”, “recommended”]	[“enak”, “banget”, “produk”, “dikirim”, “cepat”, “recommended”]
2	[“gurih”, “sekali”, “tapi”, “agak”, “mahal”]	[“gurih”, “sekali”, “tapi”, “agak”, “mahal”]

Tahap ini mengurangi jumlah kata tanpa kehilangan makna utama dari ulasan pelanggan.

Stemming

Tahap *stemming* mengubah kata menjadi bentuk dasarnya agar variasi kata memiliki representasi yang sama.

Tabel 8. Hasil *Stemming*

No	Sebelum Proses	Sesudah Proses
1	[“enak”, “banget”, “produk”, “dikirim”, “cepat”, “recommended”]	[“enak”, “banget”, “produk”, “kirim”, “cepat”, “rekomendasi”]
2	[“gurih”, “sekali”, “tapi”, “agak”, “mahal”]	[“gurih”, “sekali”, “tapi”, “agak”, “mahal”]

Dengan proses *stemming*, variasi kata seperti “dikirim” menjadi “kirim”, dan “recommended” menjadi “rekomendasi”, sehingga konsistensi data meningkat.

Normalisasi

Tahap *normalisasi* mengubah kata tidak baku atau singkatan menjadi bentuk baku agar data memiliki konsistensi dan makna yang jelas.

Tabel 9. Hasil Normalisasi

No	Sebelum Proses	Sesudah Proses
1	[“gk”, “bgt”, “produkny”, “udh”, “ok”]	[“tidak”, “banget”, “produknya”, “sudah”, “oke”]
2	[“brg”, “cepet”, “bgs”, “smoga”, “trs”]	[“barang”, “cepat”, “bagus”, “semoga”, “terus”]

Dengan proses normalisasi, variasi kata tidak baku seperti “gk” menjadi “tidak”, “bgt” menjadi “banget”, dan “brg” menjadi “barang”, sehingga teks menjadi lebih seragam dan mudah diproses pada tahap analisis berikutnya.

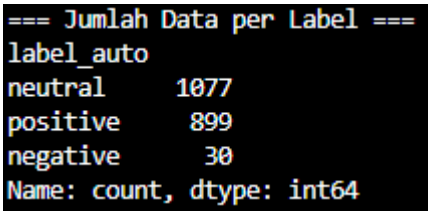
Labelling Data

Tahap selanjutnya adalah *labeling*, yaitu memberikan label kategori sentimen (*positif*, *netral*, atau *negatif*) berdasarkan konteks kalimat dan rating pelanggan.

Tabel 10. Hasil *Labelling Data*

No	Ulasan Akhir	Sesudah Proses
1	enak banget produk kirim cepat rekomendasi	Positif
2	gurih sekali tapi agak mahal	Netral
3	tidak enak dan lama sampainya	Negatif

Tahapan ini menghasilkan dataset siap latih yang akan digunakan dalam proses ekstraksi fitur menggunakan metode TF-IDF. Hasil pelabelan yang diperoleh dari distribusi sentimen sebagaimana disajikan pada gambar 4 berikut.



Gambar 4. Jumlah Data per Label

Pada gambar 4 menunjukkan bahwa ulasan dengan sentimen netral mendominasi, diikuti oleh sentimen positif, sedangkan negatif hanya menyumbang proporsi kecil. Hal ini mengindikasikan bahwa mayoritas pelanggan memberikan penilaian yang cenderung netral hingga positif terhadap produk *Nyemil.Saji*. Ulasan netral umumnya berisi komentar informatif seperti "produknya sesuai deskripsi" atau "pengiriman cepat", sementara ulasan positif mencerminkan kepuasan pelanggan terhadap rasa dan pelayanan, contohnya "enak banget, pengiriman cepat!". Sebaliknya, ulasan negatif sebagian besar menyoroti keterlambatan pengiriman atau ketidaksesuaian kemasan produk.

Ekstraksi Fitur Menggunakan TF-IDF

Tahap selanjutnya adalah proses ekstraksi fitur menggunakan metode *Term Frequency Inverse Document Frequency (TF-IDF)*. Metode ini mengubah teks hasil *preprocessing* menjadi representasi numerik yang menggambarkan tingkat kepentingan setiap kata dalam dokumen (Pratiwi et al., 2021).

Tabel 11. Hasil *Ekstraksi Fitur TF-IDF*

No	Kata	Frekuensi	Nilai TF-IDF
1	enak	124	0.876
2	cepat	98	0.851
3	rekomendasi	85	0.792
4	gurih	63	0.745
5	mahal	28	0.532

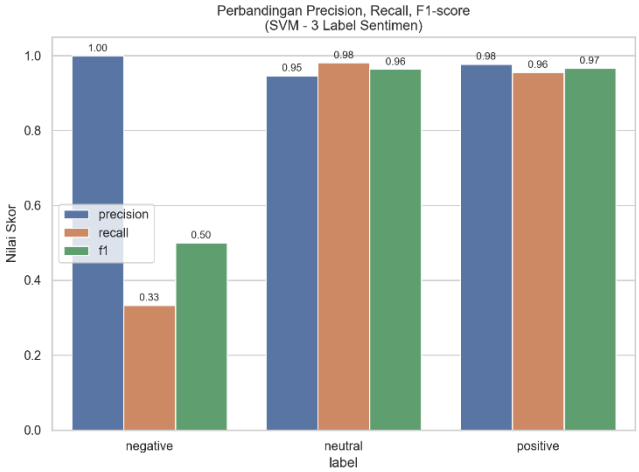
Nilai pada Tabel 11 menunjukkan kata-kata dengan bobot tertinggi yang paling berpengaruh dalam pembentukan model klasifikasi. Hal ini menegaskan bahwa ulasan pelanggan banyak menyoroti kualitas produk (*taste*) dan pelayanan (*delivery speed*), yang menjadi indikator utama kepuasan konsumen pada produk *Nyemil.Saji*.

Klasifikasi *Support Vector Machine*

Tahap berikutnya adalah klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM) dengan representasi fitur TF-IDF. Algoritma ini dipilih karena kemampuannya menangani data berdimensi tinggi dan menghasilkan batas pemisah (*hyperplane*) yang optimal antar kelas (Idris, Mustofa, & Salihi, 2023). Model dievaluasi menggunakan metrik *accuracy, precision, recall,* dan *F1-score,* sebagaimana disajikan pada Tabel 12 berikut.

Tabel 12. Hasil *Ekstraksi Fitur TF-IDF*

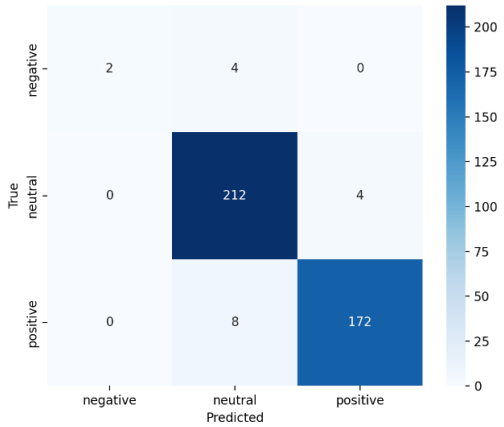
Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	1.00	0.33	0.50	6
Netral	0.95	0.98	0.96	216
Positif	0.98	0.96	0.97	180
Akurasi *(Overall)*	**0.96**			**402**



Gambar 5. Visualisasi Perbandingan *Precision, Recall, F1-Score*

Berdasarkan hasil evaluasi pada Tabel 12 dan visualisasi Gambar 5, model SVM menghasilkan akurasi sebesar 96%, yang menunjukkan tingkat ketepatan prediksi yang sangat tinggi. Nilai *precision* dan *recall* tertinggi dicapai pada kategori sentimen netral, yang berarti model mampu mengenali dan memprediksi ulasan netral secara konsisten. Nilai *F1-score* tertinggi terdapat pada sentimen positif (0.97), menandakan keseimbangan optimal antara presisi dan sensitivitas model dalam mengenali opini pelanggan yang bersifat positif.

Walaupun model menghasilkan akurasi keseluruhan sebesar 96%, nilai ini tidak sepenuhnya mencerminkan performa model yang seimbang. Akurasi tinggi terutama disumbang oleh kelas mayoritas (*netral* dan *positif*), sementara kelas minoritas (*negatif*) memberikan kontribusi kecil terhadap total prediksi benar. Artinya, meskipun model tampak sangat akurat secara umum, kemampuannya dalam mendeteksi ulasan negatif masih terbatas. Kondisi ini menunjukkan bahwa akurasi bukan satu-satunya indikator keberhasilan model, terutama ketika data tidak seimbang (Hajaroh, Suprapti, & Narasati, 2024).



Gambar 6. *Confusion Matrix*

Berdasarkan Gambar 6, sebagian besar ulasan terklasifikasi dengan benar oleh model. Sebanyak 212 ulasan netral diprediksi dengan tepat, 172 ulasan positif juga teridentifikasi sesuai label, sementara kesalahan klasifikasi minor terjadi pada beberapa ulasan negatif yang terprediksi sebagai netral. Kesalahan ini disebabkan oleh kemiripan konteks linguistik antara ulasan negatif ringan dan netral, misalnya penggunaan kata “kurang enak” atau “biasa saja” yang berada di antara dua polaritas sentimen.

Namun, performa model yang rendah pada kelas *negatif* (*recall* hanya 0,33) disebabkan oleh ketidakseimbangan jumlah data antar kelas *(imbalanced class)*. Berdasarkan hasil pelabelan, data *negatif* hanya berjumlah 30 dari total 2.006 ulasan (<2%), jauh lebih sedikit

dibandingkan dengan kelas *netral* (1.077) dan *positif* (899). Ketidakseimbangan ini menyebabkan model tidak memperoleh cukup contoh untuk "belajar" mengenali pola linguistik khas ulasan negatif, sehingga banyak ulasan negatif salah diklasifikasikan sebagai netral. Kondisi ini umum terjadi dalam *sentiment analysis* karena data pelanggan cenderung lebih banyak berisi opini positif atau netral daripada negatif (Ridho et al., 2023; Afandi et al., 2024).

Temuan ini menunjukkan bahwa model SVM memiliki sensitivitas tinggi terhadap konteks positif dan netral, namun masih perlu penguatan data latih untuk sentimen negatif yang jumlahnya sedikit.

Temuan kecenderungan sentimen *netral-positif* ini secara langsung menjawab rumusan masalah penelitian, yaitu "bagaimana pandangan pelanggan terhadap penilaian produk di Tokopedia *Nyemil.Saji*?". Berdasarkan hasil klasifikasi dan analisis TF-IDF, dapat disimpulkan bahwa pelanggan cenderung memiliki pandangan yang positif terhadap produk, dengan kepuasan terutama berfokus pada rasa produk dan kecepatan pengiriman. Dominasi kata "enak", "cepat", dan "rekomendasi" mengindikasikan dua aspek utama yang menjadi indikator kepuasan pelanggan, yakni *product quality* dan *delivery service*.

Untuk meningkatkan performa model, terutama pada kelas minoritas, penelitian selanjutnya disarankan untuk menerapkan pendekatan penyeimbangan data seperti *Synthetic Minority Over-sampling Technique* (SMOTE), *undersampling,* atau *cost-sensitive learning* agar model dapat mengenali ulasan negatif dengan lebih baik. Selain itu, eksplorasi teknik representasi fitur berbasis *word embeddings* seperti *Word2Vec* atau *BERT* juga berpotensi meningkatkan pemahaman konteks semantik antar kata, sehingga prediksi model menjadi lebih akurat dan kontekstual.



Gambar 7. *Wordcloud All Labels*

Gambar 7 menampilkan visualisasi *wordcloud* berdasarkan bobot TF-IDF dari seluruh dataset. Kata-kata dengan ukuran lebih besar memiliki nilai TF-IDF yang lebih tinggi, menandakan kontribusinya yang besar terhadap model klasifikasi. Kata seperti "enak", "kecewa", "rusak", dan "maaf" tampak dominan, yang menunjukkan bahwa ekspresi sentimen positif dan negatif sangat bergantung pada kosakata eksplisit ini. Dominasi kata "enak" menandakan banyak ulasan positif, sementara kata "kecewa" merepresentasikan ulasan negatif.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa pandangan pelanggan terhadap produk *Nyemil.saji* di Tokopedia cenderung positif, ditunjukkan oleh dominasi ulasan dengan sentimen positif dan kata-kata evaluatif seperti 'enak' dan 'renyah'. Model SVM dengan fitur TF-IDF terbukti mampu mengklasifikasikan ulasan dengan performa tinggi, terutama pada sentimen positif dan netral.

Dengan demikian, penelitian ini tidak hanya memberikan gambaran objektif tentang persepsi pelanggan terhadap produk UMKM lokal di *platform e-commerce*, tetapi juga

menegaskan pentingnya penerapan analisis sentimen berbasis *machine learning* sebagai alat bantu pengambilan keputusan strategis bagi pelaku usaha seperti *Nyemil.Saji.*

KESIMPULAN

Penelitian ini berhasil menerapkan metode *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen pelanggan terhadap produk *Nyemil.Saji* di Tokopedia. Berdasarkan hasil pengujian, model SVM dengan representasi fitur TF-IDF mencapai akurasi sebesar 96%, yang menunjukkan kinerja sangat baik dalam mengidentifikasi kategori sentimen positif, netral, dan negatif. Mayoritas ulasan pelanggan menunjukkan kecenderungan netral hingga positif, yang berarti pelanggan merasa puas terhadap kualitas produk dan kecepatan pengiriman. Meski demikian, sebagian kecil ulasan negatif masih ditemukan, umumnya berkaitan dengan keterlambatan pengiriman atau ketidaksesuaian kemasan. Temuan ini menegaskan bahwa analisis sentimen berbasis SVM dapat membantu pelaku UMKM seperti *Nyemil.Saji* untuk mengevaluasi persepsi pelanggan secara otomatis dan efisien, serta mendukung pengambilan keputusan strategis dalam peningkatan layanan dan kualitas produk di platform *e-commerce.*

Keterbatasan penelitian ini terletak pada ketidakseimbangan jumlah data antar kelas sentimen, khususnya pada kelas negatif yang jumlahnya sangat sedikit. Kondisi ini memengaruhi kemampuan model dalam mengenali pola ulasan negatif secara optimal. Oleh karena itu, pada penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data seperti *Synthetic Minority Over-sampling Technique* (SMOTE) atau *cost-sensitive learning* untuk meningkatkan performa model pada kelas minoritas. Selain itu, penelitian lanjutan juga dapat mengeksplorasi algoritma klasifikasi lain seperti *Random Forest* atau *Neural Network* untuk memberikan perbandingan kinerja yang lebih komprehensif terhadap model SVM.

REFERENSI

Afandi, A., & Ramadhani, R. (2024). *Analisis Sentimen Review Aplikasi Mobile Banking Menggunakan Metode Naïve Bayes dan Support Vector Machine*. *Jurnal Informatika dan Teknologi Digital (JITD)*, 6(2), 145–154.

Ahadiyah, R. (2024). *Analisis Sentimen Ulasan Produk di Tokopedia Menggunakan Support Vector Machine (SVM)*. *Jurnal Teknologi dan Sistem Informasi*, 8(2), 115–124.

Arsadhana, A. D., Efendi, Y., & Trihudiyatmanto, M. (2025). *Analisis Sentimen Review Film di Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine*. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 9(1), 34–45.

Ayuningrum, A. P. D., & Isa, M. (2024). *Analisis Sentimen Ulasan Produk E-Commerce Menggunakan Metode Naïve Bayes dan Support Vector Machine*. *Jurnal Sistem Informasi dan Komputer Terapan Indonesia (JSIKTI)*, 7(1), 35–42.

Clara, N., & Beni, S. (2023). *Analisis Sentimen Konsumen terhadap Layanan Marketplace Shopee Menggunakan Algoritma K-Nearest Neighbor*. *Jurnal Teknologi Informasi dan Sains Terapan (JTIST)*, 5(3), 201–210.

Daruhadi, A., & Suryawan, R. (2024). *Analisis Sentimen Pelanggan terhadap Layanan GoFood Menggunakan Support Vector Machine (SVM)*. *Jurnal Teknologi Komputer (JTK)*, 8(1), 45–53.

Fitri, S. D., Lestari, D., Bintana, R. R., Aryani, R., Ilhami, M., & Noverina, Y. (2024). *Implementasi Model Support Vector Machine dalam Analisa Sentimen Masyarakat Mengenai Kebijakan Penerapan Aplikasi MyPertamina*. *BRIDGE: Jurnal Publikasi Sistem Informasi dan Telekomunikasi*, 2(2), 176–193. https://doi.org/10.62951/bridge.v2i2.180

Hajaroh, H., Suprapti, T., & Narasati, R. (2024). *Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Produk Makanan dan Minuman di Tokopedia*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 111–118.

Harahap, M., Sihombing, B. P. A., Laia, O. A. F., Saragih, B. T., & Dharma, K. (2021). *Analisis Sentimen Review Penjualan Produk UMKM pada Kabupaten Nias dengan Komparasi Algoritma Klasifikasi Machine Learning*. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, 5(2), 147–154.

Idris, S., Mustofa, K., & Salihi, M. A. (2023). *Analisis Komparatif Kinerja Algoritma SVM dan Random Forest dalam Klasifikasi Teks Bahasa Indonesia*. *Jurnal Ilmiah Informatika dan Komputer (JIIK)*, 9(3), 211–220.

Kamal, W. W., & Ratnasari, C. I. (2020). *Analisis Sentimen Ulasan Produk: Kajian Pustaka*. *Jurnal Nasional Informatika dan Teknologi Jaringan (INFOTEKJAR)*, 6(1), 1–10.

Kosasih, R., & Alberto, A. (2021). *Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor*. *INFOTEKJAR: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(1), 1–8.

Kusuma, I. H., & Cahyono, N. (2023). *Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor*. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 8(3), 302–310.

Muzaki, A., Febriana, V., & Cholifah, W. N. (2024). *Analisis Sentimen pada Ulasan Produk di E-Commerce dengan Metode Naïve Bayes*. *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, 5(4), 758–762.

Permadi, V. A. (2020). *Analisis Sentimen Menggunakan Algoritma Naïve Bayes terhadap Review Restoran di Singapura*. *Jurnal Buana Informatika*, 11(2), 141–151.

Pratiwi, R. W., Handayani, S. F., Dairoh, D., Af'idah, D. I., Rifa, Q., & Amaliyah, G. F. (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.

Ridho, F., Nurfauziah, D., & Maulana, A. (2023). *Penerapan Analisis Sentimen Ulasan Pelanggan Menggunakan Metode Support Vector Machine (SVM)*. *Jurnal Teknologi Informasi dan Komunikasi Terapan (JATIKOM)*, 9(2), 155–162.

Siti, N., Sarjita, S., & Cahyono, E. (2024). *Analisis Sentimen Review Pelanggan pada Marketplace Tokopedia Menggunakan Support Vector Machine (SVM)*. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 13(1), 88–96.

Ulfah, N. (2023). *Analisis Sentimen pada Ulasan Produk Shopee Menggunakan Metode K-Nearest Neighbor dan Support Vector Machine*. *Jurnal Informatika dan Sains Terapan (JISTA)*, 7(2), 215–224.

Wijaya, R., Indriati, D., & Muzaki, A. (2021). *Analisis Sentimen Ulasan Produk di Marketplace Shopee Menggunakan Metode Support Vector Machine dan Naïve Bayes*. *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 7(2), 101–110.