

Analisis Sentimen Ulasan Produk Sparepart Motor Di E-Commerce Menggunakan Metode Support Vector Machine (SVM)

Reza Ega Resnanda^{1*}, Dwi Cahyono², Anik Vega Vitianingsih³

^{1,2,3} Fakultas Teknik, Teknik Informatika, Universitas Dr. Soetomo, Surabaya, Indonesia

E-mail: ^{1*}rezaega37@gmail.com, ²dwik@unitomo.ac.id, ³vega@unitomo.ac.id

(* : corresponding author)

Abstrak

Penelitian ini dilatarbelakangi oleh meningkatnya penggunaan e-commerce di Indonesia yang mendorong pentingnya analisis ulasan pelanggan sebagai dasar evaluasi kualitas produk dan layanan. Penelitian ini bertujuan untuk menganalisis sentimen ulasan produk sparepart motor merek Honda pada toko *Ducks Garage* di platform Tokopedia menggunakan algoritma Support Vector Machine (SVM). Dataset yang digunakan berjumlah 2.537 ulasan yang diperoleh melalui teknik web scraping dan diproses melalui tahapan *preprocessing* teks, meliputi pembersihan data, normalisasi, tokenisasi, penghapusan stopword, dan stemming. Pelabelan sentimen dilakukan ke dalam tiga kelas, yaitu positif, negatif, dan netral, dengan *lexicon-based* serta pembobotan fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Ketidakseimbangan distribusi data ditangani menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Model SVM diuji menggunakan tiga jenis kernel, yaitu Linear, Polynomial, dan Radial Basis Function (RBF). Hasil pengujian menunjukkan bahwa kernel RBF menghasilkan kinerja terbaik dengan akurasi 92,79%, diikuti kernel Linear 89,89%, dan kernel Polynomial 72,57%. Kesimpulan dari penelitian ini menunjukkan bahwa penerapan SVM dengan penyeimbangan data SMOTE efektif dalam mengklasifikasikan sentimen ulasan produk e-commerce dan dapat dimanfaatkan untuk mendukung pengambilan keputusan bisnis berbasis data pelanggan.

Kata kunci: E-Commerce, Analisis Sentimen, Support Vector Machine, Tokopedia

Abstract

This study was motivated by the increasing use of e-commerce in Indonesia, which highlights the importance of analyzing customer reviews as a basis for evaluating product and service quality. This study aims to analyze the sentiment of reviews of Honda motorcycle spare parts at the Ducks Garage store on the Tokopedia platform using the Support Vector Machine (SVM) algorithm. The dataset used consists of 2.537 reviews obtained through web scraping techniques and processed through text preprocessing stages, including data cleaning, normalization, tokenization, stopword removal, and stemming. Sentiment labelling was carried out into three classes, namely positive, negative, and neutral, with *lexicon-based* and feature weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) method. Data distribution imbalance was handled using the Synthetic Minority Over-Sampling Technique (SMOTE) method. The SVM model was tested using three types of kernels, namely Linear, Polynomial, and Radial Basis Function (RBF). The test results showed that the RBF kernel produced the best performance with an accuracy of 92.79%, followed by the Linear kernel at 89.89% and the Polynomial kernel at 72.57%. The conclusion of this study shows that the application of SVM with SMOTE data balancing is effective in classifying the sentiment of e-commerce product reviews and can be used to support data-driven business decisions based on customer data.

Keywords: E-Commerce, Sentiment Analysis, Support Vector Machine, Tokopedia

1. PENDAHULUAN

Perkembangan era digital di Indonesia meningkat pesat setiap tahunnya, dengan e-commerce sebagai salah satu contohnya. Pada tahun 2022, transaksi online di Indonesia mencapai 14% dari total populasi, atau setara dengan 21 juta orang [1]. Teknologi digital kini telah mengubah cara konsumen berinteraksi dengan layanan, sehingga mendorong banyak bisnis memanfaatkannya untuk menjangkau pelanggan sehingga lebih efisien. Salah satu marketplace populer di Indonesia adalah Tokopedia, yang telah merilis aplikasi *mobile* untuk platform Android. Tokopedia menjadi toko online dengan jumlah kunjungan tertinggi di Indonesia, mencapai sekitar 1,2 miliar pengunjung, terdiri atas 863,1 juta pengguna dari *mobile* dan 329,8 juta dari *desktop* [2].

Di *E-commerce* seperti Tokopedia, terdapat banyak toko yang menyediakan produk sparepart motor. Dari adanya fitur keluhan layanan dalam ulasan ini dapat mempengaruhi persepsi pelanggan dan menurunkan tingkat kepercayaan terhadap toko penyedia sparepart motor tersebut. Oleh karena itu, ulasan menjadi sumber informasi penting yang perlu dianalisis untuk memberikan solusi yang relevan untuk meningkatkan kualitas layanan dan produk sesuai dengan kebutuhan pelanggan.

Salah satu toko yang menyediakan produk sparepart motor adalah *Ducks Garage*, toko yang cukup populer di *e-commerce* tersebut. *Ducks Garage* menjadi opsi yang sangat berguna bagi individu yang membutuhkan produk sparepart motor khususnya Honda. Namun, di tengah pesatnya perkembangan teknologi digital, penting untuk tidak hanya melihat sejauh mana toko ini memenuhi kebutuhan pelanggan, tetapi juga mengevaluasi tingkat kepuasan dan tanggapan pelanggan dengan layanan yang diberikan.

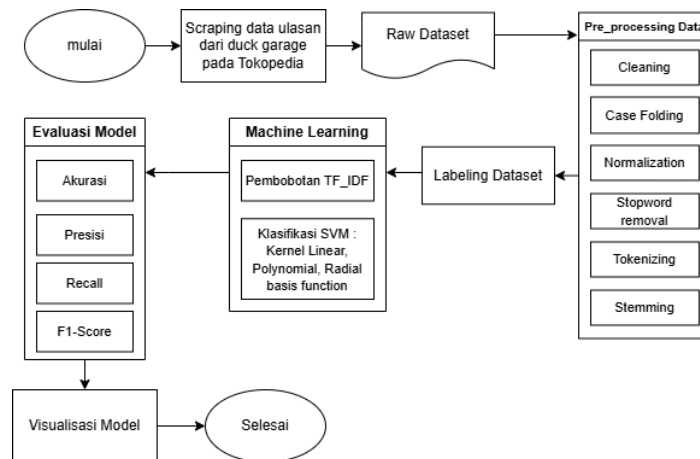
Menurut penelitian [3] yang membahas penerapan klasifikasi teks dalam analisis sentimen pada produk di Lazada yang memanfaatkan algoritma *Naive Bayes* dengan tujuan menganalisis ulasan teks pelanggan pada toko Mawar Collection. Penelitian tersebut memperoleh hasil evaluasi performa yang cukup tinggi dengan akurasi sebesar 94%, presisi dengan nilai 96%, lalu recall dengan nilai 98%, dan F1-score 97% yang mengindikasikan mayoritas ulasan pelanggan bersifat positif. Sedangkan menurut penelitian [4] yang membahas penerapan kernel pada algoritma *Support Vector Machine* (SVM) guna mengklasifikasi komisi pemberantasan korupsi pada Instagram dan Twitter. Pada riset ini dilakukan evaluasi kinerja antar model guna menilai efektivitas setiap kernel pada algoritma SVM, kernel tersebut meliputi Linear, Polynomial, dan Sigmoid dengan tujuan mengklasifikasikan komentar masyarakat di media sosial terhadap kinerja KPK. Hasil dari kernel Linear menunjukkan performa terbaik dengan akurasi 83,06%, disusul Polynomial 81,45%, dan sigmoid dengan akurasi 79,83%. Namun, kedua penelitian tersebut masih berfokus pada domain produk umum dan isu sosial, lalu belum ada kajian yang secara khusus mengangkat ulasan produk pada sektor otomotif, khususnya sparepart Honda.

Berdasarkan gap penelitian sebelumnya, penelitian ini difokuskan pada analisis kecenderungan sentimen ulasan pengguna terhadap produk sparepart motor merek Honda di toko *Ducks Garage* pada aplikasi Tokopedia dengan mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral menggunakan metode *Support Vector Machine* (SVM) dengan tiga kernel yaitu Linear, Polynomial, dan *Radial Basis Function* (RBF). SVM didefinisikan sebagai teknik klasifikasi pada bidang supervised learning di data mining, dimana metode ini unggul dalam kemampuannya memisahkan data input yang non-linear dan berdimensi tinggi [5].

Dengan demikian, penelitian ini mengkaji kecenderungan sentimen ulasan pengguna terhadap produk sparepart motor merek Honda di toko *Ducks Garage* pada aplikasi Tokopedia melalui klasifikasi sentimen positif, negatif, dan netral menggunakan metode SVM, serta berkontribusi dalam memperluas penerapan analisis sentimen pada domain otomotif khususnya sparepart Honda dan memberikan pemahaman yang lebih komprehensif mengenai persepsi pengguna terhadap kualitas produk sparepart tersebut.

2. METODE PENELITIAN

Penelitian ini dimaksudkan untuk mengevaluasi ulasan pengguna Tokopedia terhadap produk sparepart Honda di *Ducks Garage* menggunakan metode SVM. Dataset yang diperoleh untuk penelitian ini diambil menggunakan proses scraping. Setelah proses scraping, selanjutnya akan diproses dengan tahapan *cleaning data*, *case folding*, *tokenizing*, *normalization*, dan *stemming*. Setelah itu dilakukan pelabelan dataset dengan *lexicon-based*, selanjutnya berlanjut pada tahap TF-IDF yang difokuskan untuk di konversi ke dalam data teks menjadi representasi dalam bentuk angka sehingga dapat diproses lebih lanjut oleh model pembelajaran mesin. Kinerja model dinilai dengan metrik akurasi, presisi, recall, dan F1-score. Selain itu, hasil akhir penelitian disajikan dalam bentuk visualisasi grafik sebagai output dari penelitian.



Gambar 1. Research Flow

Gambar 1, menunjukkan tahapan penelitian analisis sentimen yang dimulai dari proses pengambilan data ulasan produk melalui scraping, dilanjutkan *preprocessing* data. Setelah itu, data diberi label dan diolah menggunakan metode pembobotan TF-IDF, serta algoritma SVM guna proses klasifikasi. Kinerja model akan dievaluasi melalui metrik akurasi, presisi, recall, dan F1-score, kemudian hasilnya divisualisasikan dalam bentuk grafik dan disajikan sebagai output akhir penelitian.

2.1. Pengambilan Data

Tahap pengambilan data ini melibatkan ulasan pengguna dari toko *Ducks Garage* di Tokopedia dengan metode web scraping menggunakan *library BeautifulSoup* untuk melakukan parsing terhadap struktur HTML dan XML pada Python, serta Selenium untuk mendukung proses otomatisasi dan pengujian pada platform web [6]. Data yang diambil dari platform Tokopedia berupa ulasan teks produk yang diberikan oleh pengguna dengan total 2.537 ulasan. Dataset hasil web scraping dapat diakses dan tersedia secara publik melalui GitHub [7]. Ulasan menjadi sumber utama dalam penelitian ini untuk dianalisis guna mengetahui pola sentimen terhadap produk. Metode scraping banyak digunakan dalam analisis data dan penelitian terdahulu. Metode ini dapat mengekstraksi dan memperoleh informasi yang bermanfaat untuk mengevaluasi suatu aplikasi berdasarkan opini pengguna, sehingga hasil analisis dapat digunakan sebagai dasar untuk meningkatkan kualitas dan kepuasan[8].

2.2. Preprocessing Data

Preprocessing data bertujuan menyaring dan membersihkan data ulasan pengguna dari karakter simbol, numerik, dan URL (*Uniform Resource Locator*)[9]. Beberapa tahapan *preprocessing* meliputi: pembersihan data (*Cleaning*), mengubah huruf menjadi seragam (*Case Folding*), memperbaiki ejaan tidak baku (*Normalization*), penghapusan kata umum yang tidak memiliki makna (*Stopword Removal*), memecah kalimat menjadi kata – kata terpisah (*Tokenizing*), memotong imbuhan menjadi bentuk asal (*Stemming*)[10].

2.3. Pelabelan Data

Proses pelabelan data dilakukan secara otomatis menggunakan pendekatan *lexicon-based*. Pendekatan ini menentukan orientasi sentimen suatu dokumen dengan mengakumulasi polaritas kata berdasarkan kamus yang telah dianotasi nilai orientasi semantiknya (positif, negatif, dan netral) [11]. Setiap ulasan dianalisis dengan mengidentifikasi kata bermuatan sentimen, kemudian bobot polaritasnya dijumlahkan untuk menghasilkan label sentimen secara otomatis. Proses tersebut memberikan konteks yang dibutuhkan agar model *machine learning* dapat belajar dan menghasilkan prediksi yang lebih akurat [12].

2.4. Pembobotan Kata

Data teks yang telah diberi label selanjutnya diubah ke dalam bentuk numerik untuk menghasilkan nilai bobot berdasarkan setiap kata dasar yang telah diekstraksi sebelumnya [13]. Salah satu metode yang umum digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini menghitung bobot kata dengan membandingkan frekuensi kemunculannya pada suatu dokumen tertentu serta tingkat distribusinya dikeseluruhan korpus teks [10]. Perhitungan bobot TF-IDF diformulasikan pada persamaan berikut:

$$W_{t,d} = tft,d \times idf_t \quad (1)$$

$W_{t,d}$: Bobot TF-IDF dari kata t pada dokumen d

tft,d : Jumlah kemunculan kata t dalam dokumen d

idf_t : Menunjukkan seberapa jarang kata t muncul di seluruh dokumen

2.5. Support Vector Machine (SVM)

SVM termasuk dalam teknik pemodelan klasifikasi yang bisa membedakan dua kelas yang berbeda secara tegas dan terstruktur [14]. SVM diperkenalkan dan dikembangkan oleh Vladimir Vapnik sebagai algoritma pembelajaran mesin untuk klasifikasi data, yang berfungsi untuk membentuk satu atau lebih bidang pemisah (*hyperplane*) pada ruang berdimensi tinggi, bahkan hingga ruang dengan dimensi tak terbatas [15]. Dalam SVM ada yang dinamakan margin, margin merupakan jarak pemisah antar *hyperplane* terdekat dengan titik data dari masing – masing kelas. *Hyperplane* memiliki jarak pemisah (*margin*) paling besar disebut *Optimal Hyperplane*[15]. Selain itu, SVM dapat digabungkan dengan berbagai fungsi kernel untuk mengatasi data yang tidak terpisah secara Linear. Melalui fungsi kernel, data dipetakan ke ruang fitur berdimensi lebih tinggi agar data yang sebelumnya sulit dipisahkan pada ruang asli dapat dipisahkan dengan lebih efektif [16]. Meskipun efektif, SVM masih memiliki beberapa keterbatasan pada tingkat komputasi tinggi pada kondisi jumlah fitur besar. Oleh karena itu, penerapan *preprocessing* dan *feature extraction* yang tepat guna memperoleh hasil akurasi analisis sentimen yang lebih baik [17]. Perhitungan *hyperplane* diformulasikan pada persamaan berikut:

$$(w \cdot x_i + b) \leq 1, y_i = -1 \quad (2)$$

$$(w \cdot x_i + b) \geq 1, y_i = 1 \quad (3)$$

Keterangan:

x_i : representasi data pada posisi ke- i

w : Menentukan arah dan posisi bidang pemisah (*hyperplane*).

x_i : representasi bobot kelas pada data ke- i

b : Acuan nilai bias dalam model

y_i : kelas yang terkait dengan data ke- i

1. Kernel Linear

Dalam algoritma SVM, fungsi kernel memiliki peran penting dalam membantu model menemukan batas pemisah yang optimal, terutama ketika data tidak dapat dipisahkan secara Linear. Kernel Linear dimanfaatkan ketika data dapat dipisahkan dengan garis lurus atau bidang datar [18]. Kernel Linear membantu model dalam memahami perbedaan dasar antara ulasan positif, negatif, dan netral secara cepat dan sederhana. Berikut adalah rumus fungsi Linear:

$$K(x_i, x) = x_i x \quad (4)$$

2. Kernel Polynomial

Fungsi kernel Polynomial merupakan salah satu komponen penting pada algoritma SVM. Kernel ini dimanfaatkan terutama ketika data tidak dapat dipisahkan secara Linear. Sebagai bentuk perluasan dari kernel Linear, fungsi Polynomial berperan dalam mengukur tingkat kesamaan antara vektor-vektor sampel pelatihan di dalam ruang fitur berdimensi lebih tinggi [18]. Kernel Polynomial juga berperan dalam meningkatkan kemampuan model dalam mengidentifikasi pola sentimen dalam teks, sehingga bisa meningkatkan akurasi pada data yang terpisah secara Linear. fungsi kernel Polynomial dapat dinyatakan dalam bentuk persamaan matematis berikut:

$$K(x_i, x) = (x_i x)^d \quad (5)$$

3. Kernel Radial Basis Function (RBF)

Kernel RBF memiliki tingkat fleksibilitas yang tinggi dalam mengatasi pola data yang kompleks. Selain itu, SVM dioptimalkan melalui pengaturan parameter model dengan kernel RBF, khususnya melalui parameter C dan gamma, dapat meningkatkan kinerja model dengan cara menyeimbangkan margin pemisah serta tingkat kesulitan klasifikasi [19]. Kernel RBF membantu meningkatkan kinerja SVM dalam mengenali pola non-linear pada teks, sehingga hasil klasifikasi sentimen menjadi lebih akurat dan adaptif terhadap variasi bahasa. Fungsi RBF dapat dirumuskan sebagai berikut:

$$K(x_i, x) = \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right) \quad (6)$$

2.6. Evaluasi Model

Tahap berikutnya setelah proses klasifikasi adalah evaluasi model untuk menilai tingkat akurasi dan performa algoritma yang digunakan. Evaluasi model dimanfaatkan untuk menilai kinerja dari algoritma yang sudah dibangun. Metode evaluasi yang sering diterapkan dalam menilai kinerja model klasifikasi adalah confusion matrix atau matriks kesalahan, yang berfungsi untuk mengukur kemampuan model dalam melakukan klasifikasi data secara benar melalui perbandingan antara hasil prediksi dan label sebenarnya pada data [20]. Matriks ini menampilkan jumlah data yang termasuk dalam kategori True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Keempat bagian tersebut bermanfaat untuk memberikan gambaran yang lebih mendetail mengenai performa algoritma dalam memisahkan sentimen positif, negatif, dan netral, sehingga membantu peneliti memahami tingkat akurasi, kesalahan, serta efektivitas model dalam melakukan klasifikasi data [21]. Struktur *confusion matrix* secara umum ditunjukkan dalam tabel 1.

Tabel 1. *Confusion Matrix*

	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Negatif (FN)	True Negatif (TN)

Selain itu, hasil confusion matrix memiliki kegunaan untuk menghitung beberapa metrik evaluasi penting, seperti:

1. Akurasi digunakan untuk mengukur tingkat kesesuaian antara hasil prediksi model melalui nilai sebenarnya, dengan mengukur seberapa proporsi total data yang berhasil dikelompokkan secara optimal oleh model [20]. Metrik ini umum digunakan, namun jika terdapat data yang tidak seimbang dapat menimbulkan bias dalam penilaian kinerja model. Rumus akurasi sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (7)$$

2. Presisi, adalah metrik evaluasi yang mengukur tingkat ketepatan prediksi model cara membandingkan total data relevan yang berhasil dikelompokkan secara optimal terhadap keseluruhan data yang diprediksi sebagai positif oleh model [20]. Rumus precision sebagai berikut:

$$Precision = \frac{TP}{FP+TP} \quad (8)$$

3. Recall, adalah metrik yang mengukur sejauh mana sistem berhasil menemukan seluruh data yang relevan, melalui membandingkan total data berkaitan yang berhasil teridentifikasi oleh sistem terhadap jumlah data relevan sebenarnya [20]. Rumus recall sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

4. F1 – Score, merupakan metrik yang menggabungkan kelemahan recall dan akurasi dengan menghitung rata-rata harmonik dari keduanya untuk menilai kinerja pada kelas positif. Namun, karena fokusnya hanya pada kelas positif, F1-Score belum sepenuhnya mencerminkan performa terhadap kelas negatif. Penggunaan versi berbobot dapat mengatasi hal ini dengan mempertimbangkan seluruh kelas dan distribusinya [22]. Rumus F1 – Score sebagai berikut:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{2TP}{2TP + FP + FN} \quad (10)$$

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian ini berasal dari pendekatan web scraping dari toko online *Ducks Garage* yang terdapat pada platform Tokopedia. Data yang dikumpulkan difokuskan pada produk sparepart motor dengan merek Honda, sehingga relevan dengan topik penelitian. Penelitian ini menggunakan dataset berjumlah 2.537 ulasan pengguna hasil web scraping dari toko *Ducks Garage* pada platform Tokopedia, yang dikumpulkan dengan rentang waktu Mei hingga Oktober 2025. Data tersebut merupakan hasil proses web scraping dan telah dipublikasikan secara terbuka dalam format CSV, sehingga dapat diakses dan digunakan kembali untuk keperluan validasi dan penelitian lanjutan melalui repositori GitHub[7]. Ringkasan output pengumpulan data tersebut terdapat pada tabel 2.

Tabel 2. Dataset

Full_Text
"Barangnya bagus sesuai dengan yang di aplikasi, pengiriman cepat ,"
"Barang nya bagus,tp tolong itu baud2 nya gerangan semua hrs lbh teliti dan amanah lagi"
"Bahanya bagus cetakan rapi, sudah bonus baut juga barang ori AHM. makasi duck"
"Maaf barang syaa terima pecah 1 jadi saya harus beli cermin me tukang kaca lagi, sudah cht seller tidak ada solusi, over all sebenarnya kalau ga pecah barangnya worth it banget"
"agak kecewa yang satunya dratnya malah jelek ngerusak drat lubang sepiionnya"

Setelah data diperoleh melalui web scraping, dilakukan tahap *preprocessing* untuk menghilangkan *noise* dan elemen tidak relevan. Tahap ini bertujuan mengubah data mentah menjadi terstruktur dan siap dianalisis dengan 6 langkah *preprocessing*, yakni *data cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming*. Berikut tampilan raw dataset hasil web scraping ulasan produk dari toko *Ducks Garage* di Tokopedia sebelum dilakukan proses pembersihan data.

Tabel 3. Raw Dataset

Raw Dataset
Mff baru kasih ulasan 📦 paketnya udah datang kemarin dan barang nya bagus real pick, bahannya tebal dan kokoh. Recomend bngt buat yg mau beli jangan ragu checkout

Pada tabel 3, menampilkan raw dataset hasil web scraping yang belum melalui proses pembersihan, sehingga masih mengandung kesalahan penulisan, duplikasi, karakter tidak relevan, serta elemen-elemen yang tidak diperlukan dalam analisis. Oleh karena itu, dataset ini perlu melalui tahap *preprocessing* untuk menghapus *noise* dan menstandarkan format teks agar lebih bersih, terstruktur, dan siap diolah lebih lanjut pada proses analisis selanjutnya, seperti ditunjukkan pada tabel 4.

Tabel 4. Text Preprocessing

Text Preprocessing	
<i>Cleaning data</i>	Mff baru kasih ulasan 📦 paketnya udah datang kemarin dan barang nya bagus real pick, bahannya tebal dan kokoh. Recomend bngt buat yg mau beli jangan ragu checkout
<i>Case Folding</i>	mff baru kasih ulasan paketnya udah datang kemarin dan barang nya bagus real pick, bahannya tebal dan kokoh. recomend bngt buat yg mau beli jangan ragu checkout

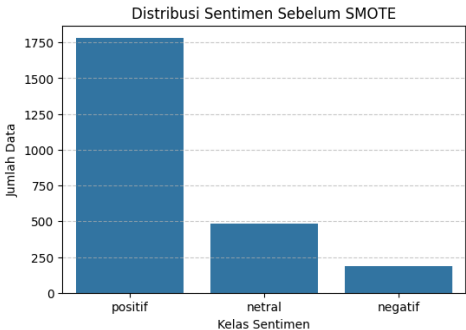
Normalisasi	maaf baru kasih ulasan paketnya udah datang kemarin dan barang nya bagus real pick, bahannya tebal dan kokoh. Rekomend banget buat yang mau beli jangan ragu check out
tokenizing	"maaf", "baru", "kasih", "ulasan", "paketnya", "udah", "datang", "kemarin", "dan", "barang", "nya", "bagus", "real", "pick", "bahannya", "tebal", "dan", "kokoh", "recomend", "banget", "buat", "yang", "mau", "beli", "jangan", "ragu", "check", "out"
Stopword removal	"kasih", "ulasan", "paketnya", "datang", "barang", "bagus", "real", "pick", "bahannya", "tebal", "kokoh", "recomend", "banget", "beli", "ragu", "check", "out"
Stemming	"kasih", "ulas", "paket", "datang", "barang", "bagus", "real", "pick", "bahan", "tebal", "kokoh", "rekomendasi", "banget", "beli", "ragu", "cek", "out"

Tahap *text preprocessing* menghasilkan data yang siap digunakan untuk proses pelabelan dengan pendekatan *lexicon-based*, di mana setiap ulasan dikategorikan sebagai positif, negatif, atau netral sesuai makna dan konteksnya. Pelabelan ini penting sebagai dasar model dalam mengenali pola emosi dan menghasilkan klasifikasi yang akurat. Hasil pelabelan diperlihatkan pada tabel 5.

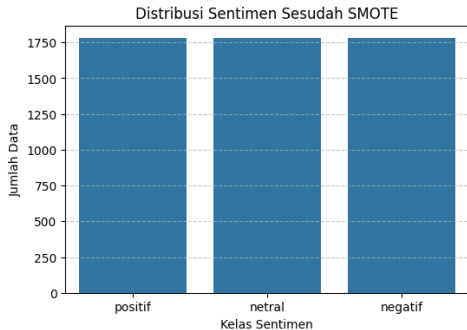
Tabel 5. Pelabelan Data

Data Stemming	Score	Label
"barang bagus sesuai yang aplikasi kirim cepat"	1	Positif
"barang datang pecah kak pack kurang bagus"	-1	Negatif
"trimakasi atas pelayananya"	0	Netral

Setelah proses pelabelan sentimen menggunakan pendekatan berbasis leksikon, tahap selanjutnya adalah memvisualisasikan distribusi ulasan berdasarkan kategori sentimen. Diagram batang digunakan untuk menggambarkan proporsi ulasan yang bersifat positif, netral, dan negatif dalam dataset. Visualisasi ini bertujuan untuk memberikan gambaran umum mengenai persebaran opini pengguna terhadap produk, sekaligus membantu dalam mengidentifikasi kecenderungan sentimen yang dominan sebelum dilakukan proses analisis lebih lanjut.



Gambar 2. Sentimen Sebelum Smote

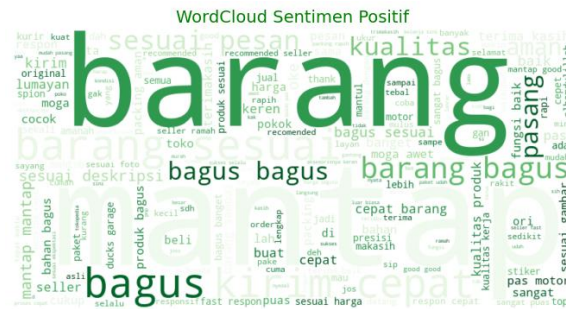


Gambar 3. Sentimen Sesudah Smote

Berdasarkan Gambar 2 dan Gambar 3, menunjukkan bahwa ada ketidakseimbangan jumlah data pada setiap kelas sentimen yang cukup signifikan. Distribusi data ulasan menunjukkan sentimen positif mendominasi dengan jumlah sekitar 1.779 ulasan, diikuti sentimen netral sekitar 486 ulasan, dan sentimen negatif yang berjumlah 188. Kondisi ini berpotensi menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan kurang optimal dalam mengenali sentimen minoritas, khususnya sentimen negatif. Dalam mengatasi permasalahan tersebut, digunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*, distribusi data menjadi lebih seimbang, dengan jumlah data pada masing-masing kelas sentimen relatif sama. Proses penyeimbangan ini diharapkan dapat meningkatkan kemampuan model dalam mempelajari pola pada setiap kelas secara lebih proporsional, sehingga performa klasifikasi menjadi lebih akurat, terutama dalam mendeteksi sentimen netral dan negatif.

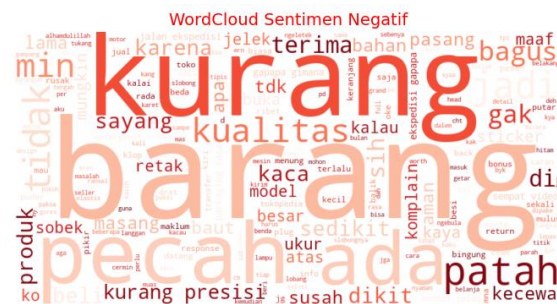
Setelah pelabelan, dibuat WordCloud untuk memvisualisasikan kata dengan frekuensi tinggi. Dalam pelabelan pada sentimen positif, terdapat beberapa kata seperti “barang”, “bagus”,

“sesuai”, serta “cepat” paling menonjol, mencerminkan kepuasan pengguna terhadap produk dan layanan, seperti yang sudah divisualisasikan pada Gambar 4.



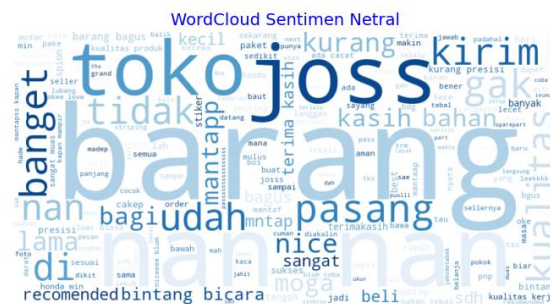
Gambar 4. Positive WordCloud

Gambar 5, memperlihatkan sentimen negatif dengan kata dominan seperti “kurang”, “patah”, “rusak”, “sobek”, dan “kecewa”, yang mencerminkan keluhan pelanggan terhadap kualitas produk.



Gambar 5. Negative WordCloud

Sedangkan Gambar 6, menunjukkan sentimen netral, kata-kata yang sering muncul seperti “barang”, “toko”, “pasang”, dan “beli” cenderung bersifat informatif tanpa mengandung emosi kuat, baik positif maupun negatif.



Gambar 6. Netral WordCloud

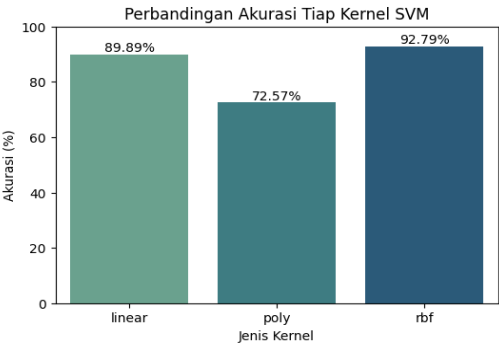
Tahapan berikutnya yaitu ekstraksi fitur yang diterapkan pada data ulasan produk sparepart motor Honda di toko *Ducks Garage* pada platform Tokopedia. Proses ini berfungsi untuk memodifikasi ulasan teks untuk menghasilkan representasi numerik agar dapat dipahami oleh model pembelajaran mesin. Melalui metode TF-IDF (*Term Frequency–Inverse Document Frequency*), pemberian bobot dilakukan pada setiap kata berdasarkan tingkat kontribusinya terhadap dokumen. Kata umum seperti “barang”, “pengiriman”, atau “produk” memiliki bobot kecil karena sering muncul di banyak ulasan, sedangkan kata seperti “tebal”, “kokoh”, atau “rusak” memiliki bobot lebih tinggi karena lebih berpengaruh terhadap sentimen pengguna. Hasil dari ekstraksi fitur ini kemudian digunakan untuk parameter awal pada model SVM guna mengidentifikasi suatu data ulasan berlabel positif, negatif, atau netral.

Pengujian model SVM dilakukan dengan menerapkan variasi 3 kernel, yaitu Linear, Polynomial, dan RBF. Setiap kernel memberikan hasil performa yang beragam, seperti terlihat pada Tabel 6.

Tabel 6. Hasil Uji Kernel

Kernel	Akurasi	Precision	Recall	F1 - Score
Linear	0.8989	0.92	0.89	0.90
Polynomial	0.7257	0.79	0.74	0.73
RBF	0.9279	0.94	0.92	0.93

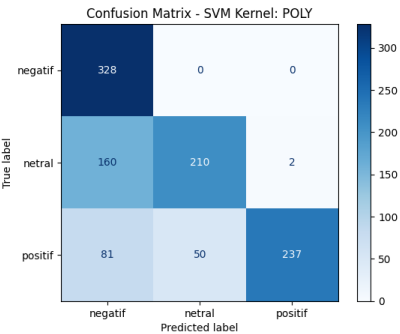
Hasil pengujian model SVM menggunakan tiga kernel menunjukkan bahwa kernel RBF memberikan kinerja paling optimal dengan nilai akurasi sebesar 92,79%, diikuti oleh kernel Linear sebesar 89,89% dan kernel Polynomial sebesar 72,57%. Selain dari sisi akurasi, kernel RBF juga memperoleh nilai presisi, recall, dan F1-score tertinggi dibandingkan kernel lainnya, yang masing-masing mencapai 0,94, 0,92, dan 0,93. Hal ini menunjukkan bahwa kernel RBF mampu mengklasifikasikan data sentimen dengan baik pada seluruh kelas. Kernel Linear juga menunjukkan performa yang cukup baik dengan nilai evaluasi yang stabil, sedangkan kernel Polynomial memiliki kinerja yang relatif lebih rendah. Secara keseluruhan, hasil pengujian memperlihatkan bahwa model SVM, khususnya dengan kernel Linear dan RBF, memiliki kinerja yang optimal dalam konteks analisis sentimen ulasan produk sparepart motor di platform Tokopedia.



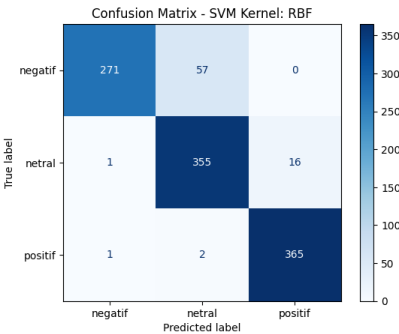
Gambar 7. Perbandingan Akurasi Kernel

Gambar 7, menunjukkan perbandingan akurasi kernel menunjukkan ketiga kernel SVM (*Linear*, *RBF*, *Polynomial*) diuji pada dataset ulasan produk. Melalui visualisasi ini, kernel dengan performa terbaik dapat diobservasi melalui nilai akurasi paling tinggi dan paling sesuai untuk digunakan pada model analisis sentimen.

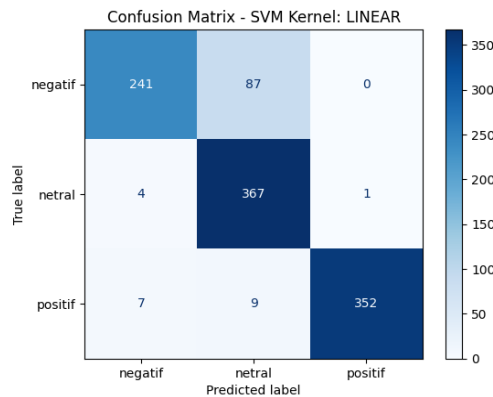
Setelah membandingkan akurasi tiap kernel, tahap selanjutnya yaitu menampilkan confusion matrix untuk melihat detail ketepatan prediksi model pada setiap kategori sentimen dapat ditunjukkan pada Gambar 8, 9, 10.



Gambar 8. Confusion Matrix Polynomial



Gambar 9. Confusion Matrix RBF



Gambar 10. Confusion Matrix Linear

4. KESIMPULAN DAN SARAN

Pada riset ini berhasil menerapkan algoritma *Support Vector Machine* (SVM) dengan memanfaatkan beberapa kernel yaitu Linear, Polynomial, dan RBF dalam analisis sentimen ulasan produk sparepart motor Honda pada toko *Ducks Garage* pada platform Tokopedia. Berdasarkan hasil pengujian dari ketiga kernel, diperoleh kernel RBF memberikan hasil terbaik dengan akurasi 92,79%, diikuti Linear memiliki akurasi 89,89%, dan terakhir kernel Polynomial yang memiliki akurasi 72,57%. Hasil ini lebih tinggi dibandingkan dengan peneliti [4] yang di mana kernel Linear menghasilkan akurasi 83.06%, Polynomial 81.45%, dan Sigmoid 79.83%. Perbedaan akurasi dipengaruhi oleh karakteristik dataset, di mana ulasan e-commerce bersifat lebih eksplisit dibandingkan komentar politik. Namun, penelitian ini memiliki keterbatasan berupa ketidakseimbangan distribusi data sentimen, di mana ulasan positif lebih dominan dibandingkan ulasan negatif dan netral, sehingga berpotensi menyebabkan model bias terhadap kelas mayoritas dan kurang optimal dalam mengenali sentimen minoritas. Secara keseluruhan, metode SVM terbukti efektif dan konsisten dalam mengklasifikasikan sentimen ulasan produk sparepart motor Honda di Tokopedia serta memiliki potensi kuat dalam mendukung pengambilan keputusan bisnis berbasis data pelanggan.

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya disarankan untuk menggunakan dataset dengan distribusi sentimen yang lebih seimbang guna meningkatkan kemampuan model dalam mengenali kelas minoritas. Selain itu, dapat dilakukan optimasi parameter pada algoritma SVM serta perbandingan dengan metode klasifikasi lain untuk memperoleh performa yang lebih optimal. Pengembangan penelitian juga dapat diarahkan pada penggunaan fitur linguistik atau pendekatan deep learning agar hasil analisis sentimen menjadi lebih akurat.

DAFTAR PUSTAKA

- [1] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.
- [2] R. Apriani and D. Gustian, "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," *Jurnal Rekayasa Teknologi Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, doi: 10.52005/rekayasa.v6i1.86.
- [3] T. K. Al Lutfani, R. Astuti, W. Prihartono, and R. Hamonangan, "Penerapan Naive Bayes Untuk Analisis Sentimen Pada Ulasan Pelanggan Di Lazada: Studi Kasus Toko Mawar Collection," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, pp.997-1003, 2025, doi: 10.23960/jitet.v13i2.6391.
- [4] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi),"

- JSTIE (Jurnal Sarjana Teknik Informatika) (E-Journal)*, vol. 9, no. 2, pp. 33-42, 2021, doi: 10.12928/jstie.v9i2.20181.
- [5] S. P. Azzahra, Y. A. Apriyanto, and A. Wijaya, "Analisis Sentimen Ulasan Aplikasi Deepl Pada Google Play Dengan Metode Support Vector Machine (SVM)," vol. 4, no. 2, pp. 59–66, 2023.
- [6] M. R. Al Hafizh, Aldi Daim Fauzan, Woro Isti Rayahu, Kiki Mustaqim, and Rahma Hanum, "Web Scraping Data Ulasan Pelanggan untuk Kemajuan Bisnis E-Commerce pada Official Store dan Non-Official Store dengan Pendekatan Natural Language Processing," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 1–9, 2025, doi: 10.47709/dsi.v5i1.5748.
- [7] E. Reza, "Dataset-sentiment-analysis." [Online]. Available: <https://github.com/rezaega/Dataset-sentiment-analysis>
- [8] S. Setyabudi and E. Aryanny, "Analisis Sentimen Penilaian Pengguna Marketplace Lazada Dengan Metode Naïve Bayes dan Support Vector Machine," vol. 10, no. 1, pp.422-433, 2025, 2025, doi: 10.35314/sww8cg21.
- [9] M. Hamka, "Analisis Sentimen Pengguna E-Commerce dan Marketplace Menggunakan Support Vector Machine," *Jurnal Rekayasa Sistem Informasi dan Teknologi*, vol 1, no. 4, pp. 273-282, 2024, doi: 10.59407/jrsit.v1i4.555.
- [10] M. Syamsul Hadi, J. Akbar, and M. F. Zulkarnain, "Analisis Sentimen Wisata Air Terjun Di Kabupaten Lombok Tengah Menggunakan Metode Support Vector Machine (SVM)," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 2, pp. 318–329, 2025, doi: 10.36080/skanika.v8i2.3578.
- [11] M. Taboada, J. Brooke, and K. Voll, "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011, [Online]. Available: <https://www.proquest.com/docview/896181231/C4F09CD9F4A6440APQ/13?accountid=13827>
- [12] R. Damanhuri and V. A. Husein, "Analisis Sentimen pada Ulasan Aplikasi Access by KAI Berbahasa Indonesia Menggunakan Word-Embedding dan Classical Machine Learning," *Jurnal Masyarakat Informatika*, vol. 15, no. 2, pp. 97–106, 2024, doi: 10.14710/jmasif.15.2.62383.
- [13] H. Harnelia, "Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (SVM)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4095.
- [14] F. Nufairi, N. Pratiwi, and F. Herlando, "Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 339–348, 2024, doi: 10.29100/jipi.v9i1.4929.
- [15] M. H. H. Aly, "Klasifikasi Diabetes Menggunakan Algoritma Support Vector Machine Radial Basis Function," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 4, no. 1, pp. 28–38, 2024, doi: 10.55606/jutiti.v4i1.3420.
- [16] S. Rabbani, *et al.*, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [17] S. D. Parameswari, *et al.*, "Studi Perbandingan Naïve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Pengguna Metaverse," *Jurnal Teknologi dan Manajemen Industri Terapan*, vol. 4, no. 3, pp. 1059–1065, 2025, doi: 10.55826/jtmit.v4i3.1122.
- [18] M. R. Pradana, W. Witanti, and A. Komarudin, "Prediksi Tingkat Keparahan Diabetes Melitus Menggunakan Support Vector Machine (SVM) dengan Kernel Polinomial dan RBF," *Jurnal Locus Penelitian dan Pengabdian*, vol. 4, no. 8, pp. 7521–7533, 2025, doi: 10.58344/locus.v4i8.4357.
- [19] M. RoisS, "Perbandingan Kinerja Support Vector Machine Dalam Klasifikasi Obesitas Dengan Pendekatan Kernel Linear Dan Radial Basis Function," *Jurnal Device*, vol. 15, no. 1, pp. 14–23, 2025.
- [20] P. Pelayanan, M. Algoritma, and N. Bayer, "Analisis Sentimen Ulasan Produk Toko Online Esrocte Untuk Peningkatan Pelayanan Menggunakan Algoritma Naive Bayes," *Blantika Multidisciplinary*, vol. 2, no. 8, pp. 667-673, 2024, doi: <https://doi.org/10.57096/blantika.v2i8.189>
- [21] N. Andrika, Analisis Sentimen Produk Fashion Lokal Pada Marketplace Dengan Klasifikasi Support Vector Machine (SVM). repository.uin-suska.ac.id, 2024. [Online]. Available: <http://repository.uin-suska.ac.id/81625/>
- [22] N. Andrika, "Analisis Sentimen Produk Fashion Lokal Pada Marketplace Dengan Klasifikasi Support Vector Machine (SVM)," Jul. 2024.