



Analisis Sentimen Ulasan *E-Commerce* Menggunakan Metode SVM

Ryzal Nur Alvandy¹, Arita Witianti²

^{1, 2}Program Studi Informatika, Universitas Mercu Buana Yogyakarta, Indonesia

Email author: Tomblok300404@gmail.com

Article Info

Article history:

Received November 18 2025

Revised November 28, 2025

Accepted Desember 10, 2025

Keywords:

e-commerce

sentiments analysis

classification

Tokopedia

SVM

TF-IDF

SMOTE

ABSTRACT

The rapid expansion of e-commerce in Indonesia has resulted in a significant rise in the number of customer reviews, which serve as a valuable source of insight for understanding consumer satisfaction. This study aims to classify or identify sentiments from product reviews on the Tokopedia platform into three categories, using the Support Vector Machine algorithm. The classification method data were ethically collected through web scraping and include review text, ratings, and the number of "likes." The preprocessing stage involved several NLP techniques such as pre-processing data representation was generated using the Term Frequency-Inverse Document Frequency method, while the issue of class imbalance was addressed using the Synthetic Minority Over-sampling Technique. Based on the test results, the SVM model achieved an accuracy of 79.48% on the test data using a linear kernel, showing the best performance in classifying positive sentiments. However, the classification of neutral and negative sentiments still requires improvement. This study demonstrates that the combination of the TF-IDF method, additional numerical features, and data balancing techniques can produce an efficient sentiment analysis model within the e-commerce domain.

Corresponding Author:

Ryzal Nur Alvandy,

Universitas Mercu Buana Yogyakarta

Jl. Jembatan Merah No. 84C, Gejayan, Yogyakarta 55283.

Email: Tomblok300404@gmail.com



1. INTRODUCTION

Seiring meningkatnya pengguna e-commerce di Indonesia mengalami lonjakan yang signifikan dalam beberapa waktu terakhir, dengan meningkatnya penetrasi internet, kemajuan teknologi informasi, dan perubahan perilaku konsumen menuju digitalisasi transaksi. Menurut data dari berbagai laporan industri, Indonesia termasuk ke dalam negara dengan perkembangan pasar e-commerce yang paling pesat di kawasan Asia Tenggara. Tokopedia, sebagai salah satu platform marketplace terbesar dan paling dominan di Indonesia, memainkan peran penting dalam mendorong transformasi digital dalam ekosistem perdagangan daring nasional. Platform ini tidak hanya menyediakan sarana bagi konsumen untuk membeli dan menjual produk, tetapi juga memungkinkan interaksi dua arah melalui fitur ulasan produk (product review) dan penilaian berupa

rating. Ulasan-ulasan ini menjadi suatu komponen utama berupa sarana untuk mencari informasi dalam menentukan pilihan konsumen dalam menentukan transaksi, serta berfungsi sebagai cerminan langsung dari kepuasan pelanggan terhadap produk maupun layanan penjual [1], [2]

Dalam konteks ini, analisis terhadap opini pelanggan menjadi suatu kebutuhan strategis, baik bagi pelaku bisnis maupun pengembang sistem informasi. Salah satu pendekatan yang relevan dan banyak digunakan untuk tujuan tersebut adalah analisis sentiment, yaitu berupa proses otomatisasi berguna mengenali serta mengelompokkan opini yang terdapat dalam suatu kelompok sentimen tertentu, yaitu diantaranya sentiment positif, netral dan negatif. Pendekatan tersebut memiliki potensi besar dalam mengukur persepsi publik secara cepat dan skalabel, serta dapat dijadikan sebagai alat pengambilan keputusan berbasis data dalam ranah manajemen pengalaman pelanggan (customer experience management) [3], [4].

Dalam implementasinya, teknik analisis sentimen sangat bergantung pada keandalan algoritma klasifikasi dan kualitas representasi fitur dari data teks yang dianalisis. Salah satu algoritma paling populer dan sering diterapkan dalam permasalahan klasifikasi sentiment yaitu Support Vector Machine atau kita menyebutnya sebagai SVM. Mempunyai keunggulan dalam mengelola atau memproses suatu data dengan dimensi yang tinggi seperti data teks, serta mampu membentuk batas keputusan untuk memisahkan kedua kategori data dengan margin tertinggi. Pada penelitian terdahulu dapat disimpulkan bahwa analisis sentimen menggunakan metode SVM menghasilkan keakuratan tinggi di berbagai skenario klasifikasi teks, baik untuk opini produk, ulasan layanan, maupun analisis media sosial [5], [6], [7].

Efektivitas algoritma SVM dapat ditingkatkan dengan pengaturan parameter kernel serta pemilihan fitur teks yang tepat. Beberapa studi melaporkan bahwa pemanfaatan teknik representasi seperti *Term Frequency–Inverse Document Frequency* mampu meningkatkan kinerja pengelompokan guna memberikan bobot penting terhadap kata-kata yang memiliki keunikan pada dokumen tertentu namun jarang muncul secara keseluruhan di seluruh kumpulan dokumen korpus [8], [9]. Pemanfaatan seperti Term Frequency–Inverse Document Frequency atau yang biasa disebut TF-IDF, Word2Vec, serta pendekatan berbasis leksikon telah menjadi strategi umum dalam representasi data teks [10], [11]. Kombinasi SVM dengan pendekatan semantik berbasis Vector Space Model (VSM) juga terbukti mampu meningkatkan akurasi klasifikasi [12].

Perbandingan performa antara SVM serta algoritma lain seperti Naive Bayes, K-Nearest Neighbor (KNN), maupun Random Forest juga telah dilakukan dalam berbagai penelitian. Hasil dari studi-studi tersebut menunjukkan bahwa tidak ada algoritma tunggal yang unggul secara universal, namun SVM sering menunjukkan performa superior dalam konteks data teks dengan preprocessing yang memadai [3], [6]. Dalam aplikasi e-commerce, klasifikasi sentimen juga mulai diperluas untuk mendeteksi ulasan palsu (*fake reviews*), yang menjadi isu penting dalam menjamin integritas sistem rekomendasi produk. Studi oleh [13] misalnya, menunjukkan bagaimana algoritma pembelajaran mesin dapat digunakan untuk mendeteksi anomali dalam pola ulasan pada platform seperti Amazon.

Di Indonesia sendiri, adopsi teknologi data mining dan machine learning mulai merambah ke berbagai sektor strategis, termasuk sektor bisnis, finansial, pendidikan, dan energi. Penelitian-penelitian lokal menunjukkan bahwa penerapan metode klasifikasi, prediksi, dan klusterisasi berbasis kecerdasan buatan semakin digunakan untuk meningkatkan efisiensi operasional dan kualitas layanan [14], [15], menandakan semakin luasnya penerapan kecerdasan buatan dalam pengambilan keputusan. Dalam penelitian ini, algoritma SVM digunakan untuk mengklasifikasikan sentimen dari ulasan pelanggan pada Tokopedia. Proses klasifikasi dilakukan dengan memanfaatkan fitur-fitur teks yang direpresentasikan menggunakan metode TF-IDF, serta informasi numerik tambahan seperti rating dan jumlah "like" pada ulasan. Selain itu, mengurangi kesenjangan kelas yang tidak seimbang, maka digunakan teknik oversampling dengan menerapkan *Synthetic Minority Over-sampling Technique*, dengan tujuan untuk menghasilkan model klasifikasi lebih adil dan akurat terhadap semua kelas.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analitik berbasis ulasan pelanggan yang lebih cerdas, responsif, dan mampu memberikan nilai tambah bagi pelaku e-commerce dalam memahami dinamika kepuasan pelanggan dengan cakupan analisis yang lebih mendalam. Selain itu, berdasarkan hasil penelitian bisa digunakan dalam pengembangan sistem rekomendasi produk, sistem umpan balik otomatis, serta analisis tren opini konsumen yang dapat mendukung strategi pemasaran digital yang lebih efektif [16].

2. METHOD

Penelitian ini menerapkan pendekatan kuantitatif dengan memanfaatkan teknik data mining serta Natural Language Processing (NLP) untuk menganalisis sentimen dari ulasan pelanggan yang tersedia di aplikasi Tokopedia. Pendekatan ini bertujuan untuk menghasilkan model klasifikasi otomatis yang dapat memetakan opini pengguna menjadi tiga kelompok sentimen. Pemilihan pendekatan kuantitatif didasarkan pada kebutuhan untuk melakukan analisis numerik terhadap sejumlah besar data tekstual, serta mengukur performa klasifikasi menggunakan metrik evaluasi yang terukur dan objektif..

2.1 Data Collection

Data yang dimanfaatkan dalam penelitian ini bersumber dari salah satu toko daring di platform Tokopedia. Pengumpulan data dilakukan secara legal dan etis menggunakan metode web scraping. Untuk keperluan ini, digunakan ekstensi Web Scraper.io yang dijalankan pada browser Google Chrome. Data yang dikumpulkan terdiri dari teks ulasan pelanggan, rating dalam bentuk bintang (1 hingga 5), serta jumlah “like” atau tanda suka yang diberikan pengguna lain terhadap ulasan.

Tabel 1. Data hasil scraping

rating	ulasan	Like
bintang 5	Hanya bintang yg bicara	Membantu
bintang 5	suka bnget bgus	Membantu
bintang 3	Paketnya baru sampe lumayanlah buat sehari hari cuman harus pake dalemn agak nerawang sich tp cukuplah dgn harga segitu	Membantu
bintang 3	Kurang puas bahannya panas	Membantu
bintang 5	Alhamdulillah ?? barang udah sampai barang aman lumayan lah dengan harga segitu... terima kasih ??	Membantu

Dapat dilihat data yang di dapat pada Tabel 1 sebagaimana mestinya proses pengumpulan data hanya mencakup informasi publik yang tersedia secara terbuka, tanpa menyentuh data pribadi atau informasi sensitif pengguna, sehingga sesuai dengan prinsip etika pengambilan data digital.

2.2 Preprocessing

Sebelum dilakukan proses klasifikasi, data teks hasil scraping terlebih dahulu diproses dalam tahap prapemrosesan (text preprocessing) untuk meningkatkan kualitas dan homogenitas data. Proses ini melibatkan sejumlah tahapan standar dalam NLP. Dimulai dengan mengubah huruf dari huruf besar menjadi huruf kecil, guna untuk menyelarakan semua huruf sehingga menghindari terjadinya penggunaan kata berlebihan yang tidak ada artinya. Selanjutnya dilakukan cleaning karakter berupa angka, simbol, dan emotikon dalam kalimat yang tidak terpakai dalam preprocessing.

Setelah itu, dilakukan tokenization untuk memecah kalimat menjadi token-token individual (kata), yang kemudian diproses dimulai dengan penghapusan kata penghubung guna menyederhanakan data agar efektif, sehingga dapat menjadikan optimal dalam proses preprosesing. Dan untuk yang terakhir dilakukan stemming kata, memakai pustaka sastrawi yang khusus dirancang untuk Bahasa Indonesia, dengan tujuan mengembalikan setiap kata ke bentuk dasarnya. Proses ini bertujuan untuk menyederhanakan variasi kata dan memperkecil dimensi fitur yang akan digunakan dalam tahap klasifikasi.

```

                                ulasan \
0          Hanya bintang yg bicara
1                                suka bnget bgus
2          Bagus kayanya agak panas dipake bahannya.
3                                Overall good dipake masih longgar..
4 Alhamdulillah barang sudah sampai barang bagus...

                                ulasan bersih
0          bintang yg bicara
1                                suka bnget bgus
2          bagus kaya panas dipake bahan
3                                overall good dipake longgar
4 alhamdulillah barang barang bagus sangat cocok

```

Gambar 1. Data hasil praproses

Berikut perbandingan data sebelum dilakukan tahapan prprocessing dan setelah melalui tahapan preprocessing dapat dilihat pada gambar 1.

2.3 Feature Extraction

Setelah melalui proses praproses, teks yang telah dibersihkan dikonversi menjadi bentuk representasi numerik dengan menerapkan Term Frequency–Inverse Document Frequency. Berguna untuk menjadikan setiap kata berbobot sesuai dengan sering dipergunakan dalam jumlah kata di dokumen tertentu, dan seberapa sedikit penggunaan kata tersebut di dokumen.

Sehingga, kata yang banyak digunakan pada satu dokumen namun jumlahnya minim ditemukan di data lain, maka diperoleh nilai optimal, menandakan bahwa kata itu mempunyai nilai yang berbeda tapi signifikan terhadap dokumen tersebut. Berikut rumus yang digunakan dalam TD-IDF:

$$TF - IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right)$$

dimana:

- t / term (kata),
- d / dokumen (ulasan),
- TF(t,d) adalah jumlah kemunculan t di dokumen d,
- DF(t) adalah frekuensi dokumen yang mengandung term t,
- N adalah total frekuensi dokumen dalam korpus.

2.4 Handling Imbalanced Data

Ketidakseimbangan jumlah data antar kelas, terutama dominasi ulasan dengan sentimen positif dalam e-commerce merupakan permasalahan yang kerap muncul di proses pengelompokan sentimen. Untuk menanggapi hal tersebut, Synthetic Minority Over-Sampling Technique merupakan metode yang tepat untuk digunakan sebagai penghasil sampel-sampel sintesis pada kelompok yang jarang muncul yaitu dengan cara melakukan interpolasi antar titik data di ruang vektor fitur.

Proses ini memungkinkan perluasan kelas minoritas tanpa melakukan duplikasi langsung dari data yang ada, sehingga membantu model untuk belajar dengan lebih baik terhadap karakteristik dari ketiga kelas sentimen secara seimbang.

2.5 Classification Method

Algoritma **Support Vector Machine** dengan kernel linear merupakan model klasifikasi dalam metode penelitian ini. Metode tersebut digunakan sebab mempunyai kemampuan yang optimal dalam mengatasi data berdimensi tinggi dan ketahanannya terhadap overfitting, khususnya dalam ruang vektor teks yang memiliki banyak fitur. Kernel linear digunakan karena data teks cenderung linear separable, sehingga tidak memerlukan

transformasi non-linear tambahan yang kompleks. SVM akan menentukan pengelompokan data yang efektif sehingga mampu membagi di tiga kategori sentimen dengan jarak margin yang maksimal.

Seluruh tahapan pelatihan dan pengujian model dilaksanakan menggunakan pendekatan *train-test split*, di mana proses pelatihan serta pengujian model dilakukan secara stratified untuk memastikan proporsi kelas tetap seimbang. Untuk menjamin validitas hasil, juga dilakukan evaluasi menggunakan teknik cross-validation.

2.6 Evaluation Metrics

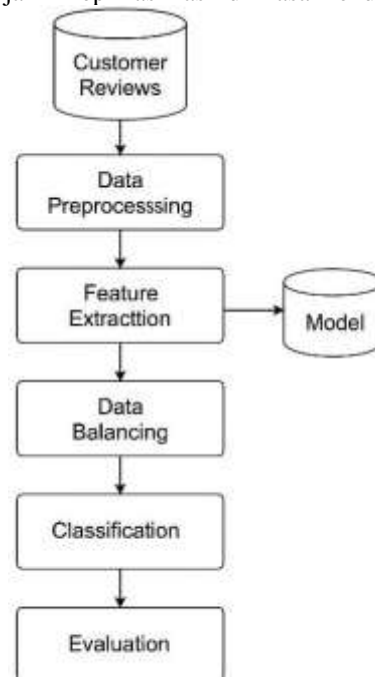
Empat metrik yang digunakan untuk evaluasi performa model diantaranya yaitu, akurasi digunakan menilai tingkat kesesuaian hasil dari presisi sesuai label asli, sedangkan presisi menggambarkan proporsi prediksi positif yang akurat.

Untuk mengetahui proses seberapa mampu model mendeteksi data yang ada termasuk dalam kelas positif untuk menangkap seluruh data positif yang tersedia disebut recall. Dan untuk F1-Score mempresentasikan keseimbangan nilai presisi serta nilai recall, keempat metrik tersebut dihitung berdasarkan confusion matrix yang menggambarkan detail klasifikasi antar kelas secara menyeluruh

2.7 System Implementation

Seluruh proses implementasi dilakukan ditulis dengan pemrograman bahasa python, juga memanfaatkan berbagai modul populer seperti scikit-learn sebagai klasifikasi, evaluasi, dan Sastrawi sebagai proses stemming Bahasa Indonesia, juga imbalanced-learn untuk implementasi SMOTE. Sistem dibangun dan dieksekusi dalam lingkungan pengembangan Jupyter Notebook yang memungkinkan integrasi antara kode program, analisis statistik, dan visualisasi hasil secara bersamaan.

Pelaksanaan penelitian dilakukan sepenuhnya secara daring, dengan pengumpulan data dilakukan dari rumah, serta pemrosesan dan pelatihan model menggunakan komputer pribadi dengan spesifikasi menengah (prosesor Intel i5 dan RAM 8GB). Semua eksperimen dan analisis dilakukan secara independen dan didokumentasikan dengan cermat untuk menjamin replikasi hasil di masa mendatang.



Gambar 2. Diagram alur sentimen

Seluruh proses dalam sistem ini divisualisasikan melalui diagram alur yang ditampilkan pada Gambar 2. Diagram tersebut menunjukkan urutan tahapan yang mencakup pengumpulan data, prapemrosesan, ekstraksi fitur, keseimbangan data, pelatihan model, sampai tahap evaluasi hasil klasifikasi.

Penelitian ini dilaksanakan secara daring, dengan proses pengumpulan data dilakukan dari rumah dan pemrosesan dilakukan menggunakan perangkat komputer dengan spesifikasi menengah. Semua eksperimen dijalankan menggunakan bahasa pemrograman Python dan pustaka-pustaka populer dalam bidang NLP dan machine learning seperti scikit-learn, Sastrawi, dan imbalanced-learn.

3. RESULT DAN ANALISIS

Tahap pertama dalam proses analisis adalah melakukan pra-pemrosesan data, yang mencakup kegiatan pembersihan teks, penghapusan karakter non-alfabet, penghilangan stopwords kecuali kata penting seperti "tidak", "kurang", dan "jangan", serta stemming menggunakan Sastrawi. Setelah tahap preprocessing, dilakukan pelabelan sentimen ditentukan berdasarkan rating, di mana ulasan dengan rating tinggi 4 dan 5(bagus), rating sedang 3 (biasa), sedangkan rating rendah 2 dan 1(jelek).

Proses pelatihan model dilakukan dengan pembagian data perbandingannya data latih(80) dengan data uji(20), sehingga mengubah data text menjadi data representasi vektor menggunakan TD-IDF, sedangkan fitur numerik (rating dan jumlah like) dinormalisasi menggunakan Standard Scaler. Untuk menangani kesenjangan frekuensi data kelompok sentimen, maka digunakan Synthetic Minority Over-sampling Technique, agar distribusi antar kelas menjadi lebih seimbang. Model SVM dengan kernel linear kemudian dilatih menggunakan data hasil SMOTE.

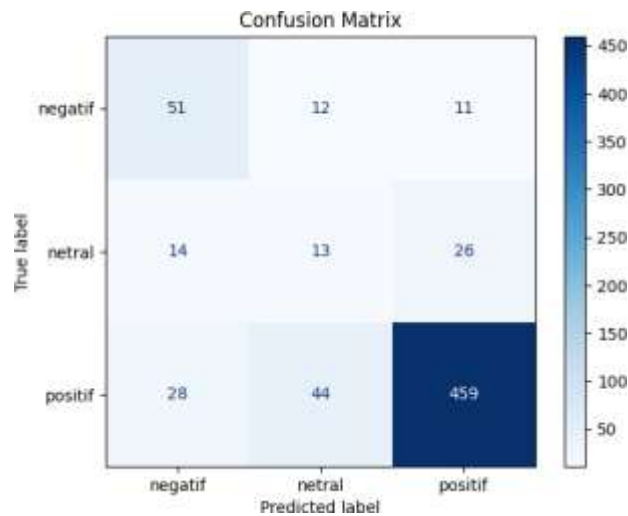
Berdasarkan hasil evaluasi, model tersebut mencapai tingkat akurasi sebesar 90,69% pada data pelatihan, dan 79,48% pada data pengujian. Meskipun terdapat sedikit penurunan akurasi pada data pengujian, hal ini menunjukkan performa yang cukup baik serta kemampuan generalisasi model yang memadai. Adapun detail evaluasi performa model pada data uji ditunjukkan melalui metrik klasifikasi berikut ini:

Tabel 2.Metrik Klasifikasi

Sentimen	Precision	Recal	F1-Score	Support
Negatif	0.55	0.69	0.61	74
Netral	0.19	0.25	0.21	53
Positif	0.93	0.86	0.89	531

Tingkat keakuratan seluruhnya mencapai **79,48%**, dengan *macro average* F1-score sebesar **57%**, dan *weighted average* sebesar **81%**, menunjukkan bahwa model memiliki kinerja terbaik dalam mengenali sentimen positif, namun kurang optimal dalam mengklasifikasikan atau membedakan antara sentimen netral dan negatif. Hal ini dapat disebabkan oleh distribusi data yang sangat timpang, di mana jumlah ulasan positif jauh lebih banyak dibanding ulasan netral dan negatif.

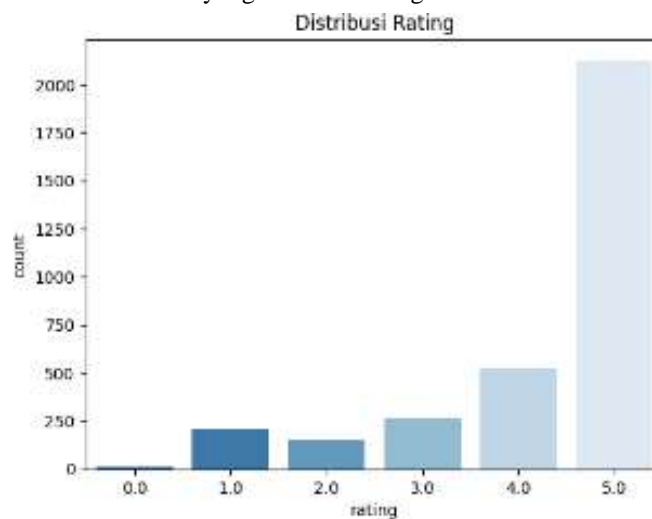
Untuk melihat sebaran prediksi model terhadap masing-masing kelas sentimen, confusion matrix digunakan sebagaimana ditampilkan pada Gambar 3.



Gambar 3 Confusion Matrix

Berdasarkan confusion matrix, dapat diketahui bahwa model mampu mengenali sentimen positif dengan sangat baik (531 benar dari total 617 prediksi positif), namun masih sering salah dalam membedakan ulasan netral dan negatif. Sebagian besar ulasan netral terklasifikasi sebagai positif, sehingga perlu dilakukan peningkatan pada fitur yang dapat membedakan kedua jenis sentimen tersebut secara lebih akurat.

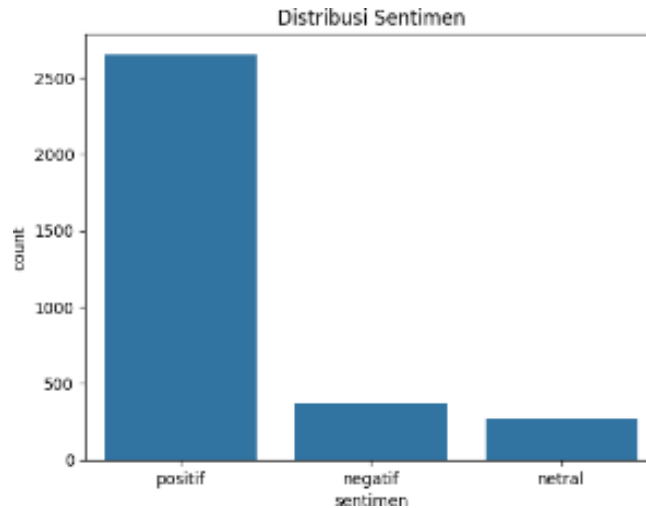
Visualisasi distribusi rating terhadap sentimen ditunjukkan pada Gambar 3. Grafik ini memperlihatkan bahwa mayoritas ulasan diberikan dengan rating bintang 5, yang secara otomatis dikategorikan sebagai sentimen positif. Sementara itu, ulasan dengan rating bintang 1 hingga 3 jauh lebih sedikit, sehingga menghasilkan distribusi data yang tidak seimbang.



Gambar 4. Distribusi Rating Ulasan Pelanggan pada Tokopedia

Gambar 4 memperlihatkan distribusi rating yang diberikan oleh pelanggan terhadap produk yang dibeli melalui salah satu toko daring di platform Tokopedia. Rentang nilai 1 sampai 5, dengan nilai yang tinggi maka menunjukkan hasil kepuasan konsumen yang optimal. Terlihat bahwa rating 5 memiliki frekuensi yang sangat dominan, yaitu lebih dari 2000 ulasan, sedangkan rating di bawahnya menunjukkan jumlah yang jauh lebih rendah. Rating 1 hingga 3 masing-masing memiliki jumlah ulasan yang relatif kecil, dengan penurunan tajam pada rating 0 dan 1.

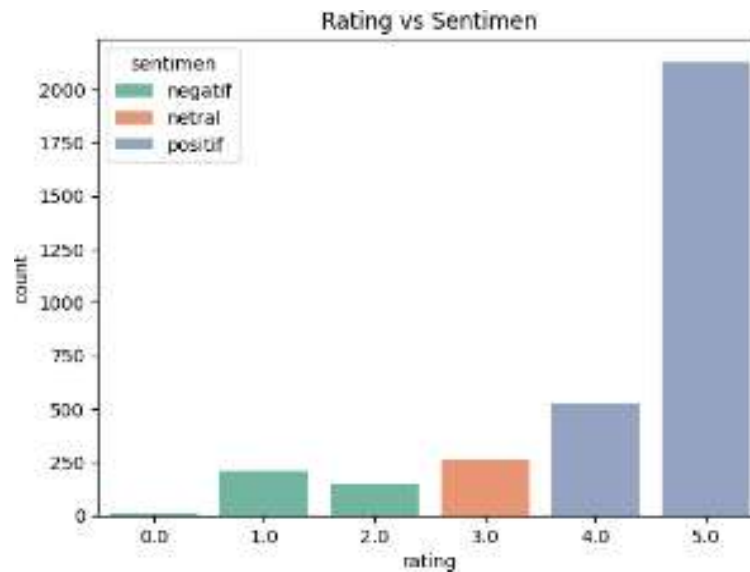
Distribusi ini menunjukkan adanya ketidakseimbangan data, di mana ulasan dengan rating tinggi (khususnya rating 5) jauh lebih banyak dibandingkan dengan ulasan ber-rating rendah. Ketidakseimbangan ini berpotensi memengaruhi performa model klasifikasi sentimen apabila tidak ditangani dengan teknik tertentu, seperti SMOTE. Oleh karena itu, penting untuk memahami karakteristik distribusi ini sebelum memasuki tahap pelatihan model, agar hasil klasifikasi tidak bias terhadap mayoritas kelas.



Gambar 5 Distribusi sentiment Ulasan Pelanggan

Gambar 5 menunjukkan distribusi kelas sentimen dari hasil labeling terhadap ulasan pelanggan yang diperoleh dari Tokopedia. Ulasan dibagi menjadi tiga kategori sentimen, yakni positif, negatif, dan netral, berdasarkan nilai rating dan analisis isi teks. Dari grafik terlihat bahwa mayoritas ulasan masuk ke dalam kategori positif, dengan jumlah melebihi 2700 data. Sementara itu, ulasan dengan sentimen negatif dan netral memiliki distribusi yang jauh lebih sedikit dan relatif seimbang satu sama lain, namun masih sangat kecil dibandingkan jumlah ulasan positif.

Distribusi ini mengindikasikan adanya ketidakseimbangan data antar kelas sentimen yang cukup signifikan. Jika tidak ditangani dengan tepat, ketidakseimbangan ini dapat menyebabkan model klasifikasi cenderung memprediksi kelas mayoritas (positif) secara berlebihan, dan mengabaikan kelas minoritas (negatif dan netral). Oleh karena itu, pada tahap pra-proses data, diterapkan teknik penyeimbangan data menerapkan SMOTE (Synthetic Minority Over-sampling Technique) guna memperbaiki ketidakseimbangan distribusi data tersebut, sehingga model yang dilatih dapat mengenali pola dari semua kelas secara adil dan akurat.



Gambar 6. Distribusi rating dan Sentimen

Gambar 6 menggambarkan hubungan antara nilai rating yang diberikan pelanggan (skala 0–5) dengan kategori sentimen (negatif, netral, dan positif) yang dihasilkan dari proses anotasi dan analisis.

Setiap batang pada grafik diwarnai sesuai dengan label sentimen yang diasosiasikan terhadap rating tersebut. Dari grafik ini, terlihat bahwa rating 5 secara konsisten berasosiasi dengan sentimen positif, dan merupakan yang paling dominan dari seluruh data, dengan lebih dari 2000 ulasan. Sementara itu, rating 1 dan 2 mayoritas diidentifikasi sebagai sentimen negatif, sedangkan rating 3 dan 4 lebih banyak diklasifikasikan ke dalam kategori netral dan sebagian kecil ke positif.

Distribusi ini mendukung validitas pendekatan anotasi yang digunakan dalam penelitian, di mana nilai rating menjadi salah satu acuan dalam menentukan label sentimen awal. Selain itu, grafik tersebut memberikan hasil relasi rating tinggi sebanding dengan kecenderungan sentiment positif, yang umum terjadi pada sistem ulasan konsumen. Informasi ini penting tidak hanya dalam proses validasi data, tetapi juga dalam memahami pola perilaku pengguna dan kualitas persepsi terhadap produk yang ditawarkan oleh penjual di platform e-commerce.

Selanjutnya, dilakukan visualisasi word cloud guna mengetahui kata yang sering digunakan pada penilaian yang tidak memuaskan. Hasilnya dapat dilihat pada Gambar 6. Kata-kata dominan yang muncul antara lain "bahan", "panas", "warna", "tipis", dan "kecewa", yang mencerminkan ketidakpuasan pelanggan terhadap kualitas material produk serta perbedaan antara ekspektasi dan kenyataan barang yang diterima.

Ketidakeimbangan data menjadi salah satu faktor yang memengaruhi hasil ini, meskipun telah dilakukan oversampling menggunakan metode SMOTE.

Oleh karena itu, disarankan untuk memperkaya jumlah data pada kelas sentimen netral serta mengeksplorasi fitur tambahan seperti analisis emosi atau penggunaan word embedding agar model dapat memahami konteks secara lebih dalam dan meningkatkan akurasi klasifikasi secara menyeluruh

ACKNOWLEDGEMENTS

Penulis menyampaikan rasa syukur dan terima kasih yang kepada dosen pembimbing atas segala bimbingan, arahan, dan saran yang begitu berharga selama berlangsungnya proses pelaksanaan penelitian ini. Ucapan terima kasih juga penulis tujukan kepada keluarga serta rekan-rekan memberikan dukungan berupa moral, motivasi, serta bantuan dalam penyusunan serta penyelesaian penelitian.

Tidak lupa, penulis menghargai seluruh yang berkontribusi diberbagai pihak yang telah memberikan bantuan, dalam proses penyelesaian tugas akhir penelitian, baik berupa partisipasi langsung maupun dukungan tidak langsung. Dengan hasil penelitian ini diharapkan dapat berperan serta mendukung dan berkontribusi sebagai pengembangan ilmu perngetahuan, serta menjadi referensi bagi penelitian selanjutnya.

REFERENCES

- [1] Ashfa, M. R., & Ishak, A. (2023). Pengaruh kualitas layanan dan kepuasan pelanggan terhadap niat beli dan pembelian sebenarnya Tokopedia. *Seleksi Manajemen: Jurnal Mahasiswa Bisnis dan Manajemen*, 2(5), Article 5.
- [2] Ganguly, B., Sengupta, P., & Biswas, B. (2024). What are the significant determinants of helpfulness of online review? An exploration across product-types. *Journal of Retailing and Consumer Services*, 78, 103748. <https://doi.org/10.1016/j.jretconser.2024.103748>
- [3] Muzaki, A., Febriana, V., & Cholifah, W. N. (2024). Analisis sentimen pada ulasan produk di e-commerce dengan metode Naive Bayes. *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, 5(4), Article 4. <https://doi.org/10.30998/jrami.v5i4.9647>
- [4] D. F. Sebastian, H. Sulistiani, and A. R. Isnain, "SENTIMENT ANALYSIS OF PUBLIC OPINION ON THE RIGHT OF INQUIRY IN INDONESIA IN 2024 USING THE SUPPORT VECTOR MACHINE (SVM) METHOD," *J. Tek. Inform. Jutif*, vol. 5, no. 4, Art. no. 4, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.1968.
- [5] Prabowo, W. A., & Azizah, F. (2020). Sentiment analysis for detecting cyberbullying using TF-IDF and SVM. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(6). <https://doi.org/10.29207/resti.v4i6.2753>.
- [6] Khan, T. A., Sadiq, R., Shahid, Z., Alam, M. M., & Su'ud, M. B. M. (2024). Sentiment analysis using support vector machine and random forest. *Journal of Informatics and Web Engineering*, 3(1), Article 1. <https://doi.org/10.33093/jiwe.2024.3.1.5>
- [7] Loomba, S., Dave, M., Arolkar, H., & Sharma, S. (2024). Sentiment analysis using dictionary-based lexicon approach: Analysis on the opinion of Indian community for the topic of cryptocurrency. *Annals of Data Science*, 11(6), 2019–2034. <https://doi.org/10.1007/s40745-023-00496-y>
- [8] Abdul Fadlil, A. F., Riadi, I., & Andrianto, F. (2024). Improving sentiment analysis in digital marketplaces through SVM kernel fine-tuning. *International Journal of Computing and Digital Systems*, 15(1), 159–171. <https://doi.org/10.12785/ijcds/160113>
- [9] Kumaragurubaran, T., Senthil Pandi, S., Naresh, G., & Ragavender, T. S. (2024, May). Navigating public opinion: Enhancing sentiment analysis on social media with CNN and SVM. In *2024 5th International Conference for Emerging Technology (INCET)* (pp. 1–5). IEEE. <https://doi.org/10.1109/INCET61516.2024.10592999>
- [10] Nurhaliza Agustina, C. A., Novita, R., Mustakim, & Rozanda, N. E. (2024). The implementation of TF-IDF and Word2Vec on booster vaccine sentiment analysis using support vector machine

- algorithm. *Procedia Computer Science*, 234, 156–163. <https://doi.org/10.1016/j.procs.2024.02.162>
- [11] Orebi, S. M., & Naser, A. M. (2024). Opinion mining in text short by using word embedding and deep learning. *Journal of Applied Data Science*, 6(1), 526–636. <https://doi.org/10.47738/jads.v6i1.438>.
- [12] Suresh Kumar, K., et al. (2024). Sentiment analysis of short texts using SVMs and VSMs-based multiclass semantic classification. *Applied Artificial Intelligence*, 38(1), 2321555. <https://doi.org/10.1080/08839514.2024.2321555>
- [13] Tabany, M., & Gueffal, M. (2024). Sentiment analysis and fake Amazon reviews classification using SVM supervised machine learning model. *Journal of Advanced Information Technology*, 15(1), 49–58. <https://doi.org/10.12720/jait.15.1.49-58>
- [14] Athallah, M. R., & Rozi, A. F. (2022). Implementasi data mining untuk prediksi peramalan penjualan produk Hj Karpel menggunakan metode linear regression. *Jurnal Sains dan Teknologi (JSIT)*, 2(3), Article 3. <https://doi.org/10.47233/jsit.v2i3.55>.
- [15] Cahyo, A. D., Anardani, S., & Yuda, Y. P. (2024). Prediksi beban daya listrik (W) menggunakan metode support vector machine (SVM) pada listrik token rumah tangga. *Jurnal Sains dan Teknologi (JSIT)*, 4(1), Article 1. <https://doi.org/10.47233/jsit.v4i1.1543>
- [16] Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 425–478. <https://doi.org/10.2307/30036540>