



Sentiment Analysis of Shopee Product Reviews on the Instagram Application Using the K-Nearest Neighbors Algorithm

Analisis Sentimen Ulasan Produk Shopee di Aplikasi Instagram Menggunakan Algoritma K-Nearest Neighbors

Sri Lestari^{1*}, Syafira Febrianti²

^{1,2}Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Daerah Khusus Ibukota Jakarta, Indonesia.

E-Mail: ¹sri.lestari1203@gmail.com, ²syafirafebrianti0220@gmail.com

Received Aug 01st 2024; Revised Aug 11th 2025; Accepted Aug 24th 2025; Available Online Oct 30th 2025

Corresponding Author: Sri Lestari

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This study examines various consumer responses to products sold on Shopee and promoted through Instagram, categorized into positive and negative sentiments. The main issue addressed is how the public perceives these products based on reviews posted on Instagram. The research employs sentiment analysis using the K-Nearest Neighbors (K-NN) algorithm and the CRISP-DM approach to ensure a systematic and structured analysis process. The dataset consists of a collection of reviews gathered through crawling techniques using automated data retrieval. The results indicate that the K-NN algorithm can classify sentiments with an accuracy of 82.80%, a positive sentiment precision of 83.54%, and a negative sentiment precision of 23.08%. These findings suggest that the majority of consumer reviews are positive, reflecting satisfaction with product quality, price, and service, although some negative reviews highlight certain shortcomings. The proposed solution in this study is the application of machine learning techniques to automate the sentiment classification process for large and unstructured review datasets. This research is expected to provide valuable insights for businesses and e-commerce platform managers to better understand consumer perceptions and use them as a basis for decision-making in marketing strategies and service quality improvements. In conclusion, the K-NN algorithm is effective for sentiment analysis of Shopee product reviews on Instagram, with results indicating a predominance of positive sentiment.

Keywords: CRISP-DM, K-Nearest Neighbors, Sentiment Analysis, Shoppe Product

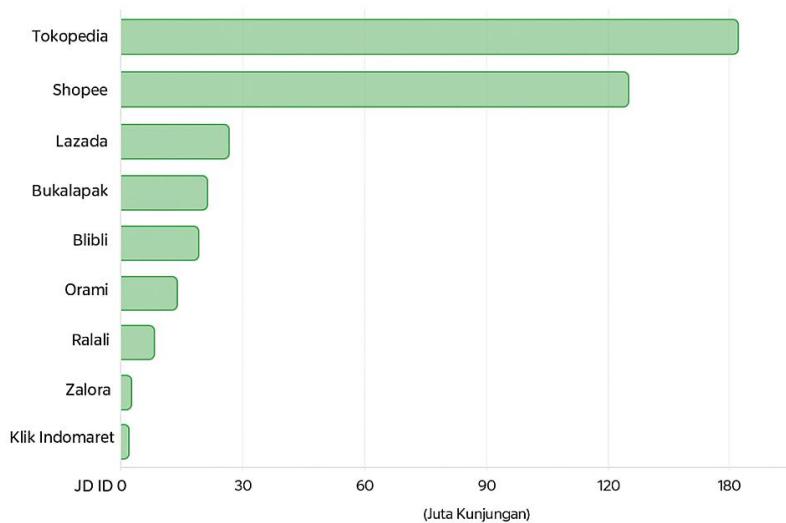
Abstrak

Penelitian ini mengkaji berbagai tanggapan konsumen terhadap produk yang dijual di Shopee dan dipromosikan melalui Instagram, yang dikategorikan ke dalam sentimen positif dan negatif. Permasalahan yang diangkat adalah bagaimana persepsi publik terhadap produk-produk tersebut berdasarkan ulasan yang diunggah di media sosial Instagram. Metode yang digunakan adalah analisis sentimen dengan algoritma K-Nearest Neighbors (K-NN) serta pendekatan CRISP-DM untuk memastikan proses analisis berjalan secara sistematis dan terstruktur. Data penelitian terdiri dari sejumlah ulasan yang dikumpulkan melalui proses *crawling* menggunakan teknik pengambilan data otomatis. Hasil penelitian menunjukkan bahwa algoritma K-NN mampu mengklasifikasikan sentimen dengan akurasi 82,80%, presisi sentimen positif sebesar 83,54%, dan presisi sentimen negatif sebesar 23,08%. Temuan ini menandakan bahwa mayoritas ulasan konsumen bersifat positif, mencerminkan kepuasan terhadap kualitas produk, harga, dan pelayanan, meskipun masih terdapat ulasan negatif yang menyoroti kekurangan tertentu. Solusi yang ditawarkan dalam penelitian ini adalah penerapan teknik *machine learning* untuk mengotomatisasi proses klasifikasi sentimen pada data ulasan yang besar dan tidak terstruktur. Penelitian ini diharapkan dapat memberikan wawasan berharga bagi pelaku bisnis dan pengelola platform *e-commerce* untuk memahami persepsi konsumen dan menggunakan sebagai dasar pengambilan keputusan dalam strategi pemasaran dan peningkatan kualitas layanan. Kesimpulannya, algoritma K-NN efektif digunakan untuk analisis sentimen ulasan produk Shopee di Instagram, dengan hasil yang menunjukkan dominasi sentimen positif.

Kata Kunci: Analisis Sentimen, CRISP-DM, K-Nearest Neighbors, Produk Shoppe

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan signifikan pada perilaku belanja masyarakat. Salah satu bentuk transformasi ini adalah meningkatnya penggunaan platform *e-commerce* untuk memenuhi kebutuhan sehari-hari [1]. Shopee, sebagai salah satu platform *e-commerce* terbesar di Indonesia, menawarkan berbagai jenis produk dengan harga kompetitif, sistem pembayaran yang beragam, serta layanan pengiriman yang luas. Popularitas Shopee tidak hanya tercermin dari tingginya jumlah transaksi, tetapi juga dari banyaknya interaksi pengguna di media sosial, termasuk Instagram. Shopee memanfaatkan media sosial seperti Instagram untuk memperluas jangkauan pemasaran dan membangun interaksi dengan konsumen [2]. Strategi ini mampu meningkatkan brand *awareness* dan mendorong minat pembelian. Meskipun popularitas Shopee cukup tinggi, masih terdapat kendala yang sering dihadapi konsumen. Beberapa di antaranya adalah ketidaksesuaian produk dengan deskripsi, keterlambatan pengiriman, dan kualitas layanan pelanggan yang bervariasi. Selain itu, ulasan konsumen yang tersebar di Instagram sering kali menunjukkan perbedaan persepsi, ada yang memberikan penilaian positif, namun tidak sedikit yang menyampaikan keluhan atau kritik. Perbedaan sentimen ini dapat memengaruhi kepercayaan dan keputusan calon pembeli [3].



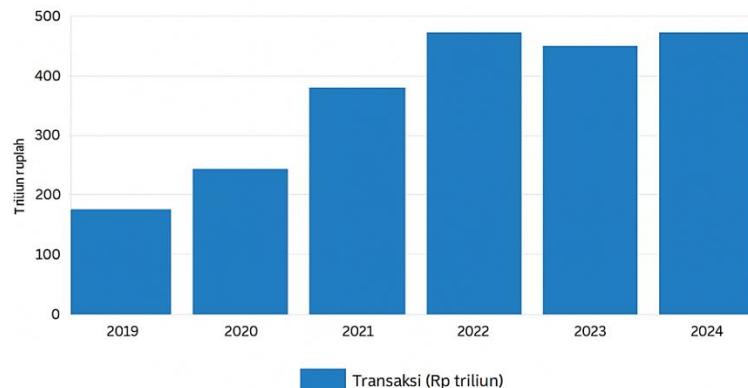
Gambar 1. Sepuluh *E-Commerce* Dengan Kunjungan Terbanyak 2022

Pada Gambar 1 dan berdasarkan informasi Sekretariat Jendral Kementerian Perdagangan Pada kuartal II tahun 2022, Tokopedia menjadi *e-commerce* yang paling banyak diakses, dengan jumlah kunjungan ke website mencapai 158,35 juta di Indonesia. Shopee berada di urutan kedua dengan 131,3 juta kunjungan, disusul Lazada dengan sekitar 26,64 juta kunjungan [4]. Masifnya penggunaan *e-commerce* ini disebabkan beberapa faktor, seperti inovasi dalam layanan dan fitur, meningkatkan promosi, penguatan ekosistem *offline* dan *online*, serta kemitraan strategis. Survey yang dilakukan Statista pada Januari 2023 terhadap 1.434 responden berusia lebih dari 18 tahun menunjukkan bahwa 70,13% responden menggunakan *e-commerce* untuk membeli barang *fashion* seperti pakaian dan sepatu. Sementara itu, 49,73% responden memilih membeli produk kecantikan dan 40,8% untuk membeli makanan pada kebutuhan konsumsi sehari-hari [5].

Terdapat pula data mengenai jumlah transaksi menggunakan *e-commerce* selama lima tahun terakhir yaitu 2019-2024. Dalam lima tahun terakhir, transaksi *e-commerce* di Indonesia mengalami pertumbuhan yang signifikan. Berdasarkan data dari Bank Indonesia, nilai transaksi *e-commerce* terus meningkat dari Rp 205,5 triliun pada 2019 menjadi Rp 487,01 triliun pada 2024 [6]. Dari data ini, terlihat bahwa lonjakan terbesar terjadi pada tahun 2021, dengan peningkatan lebih dari 50% dibanding tahun sebelumnya. Hal ini kemungkinan besar didorong oleh perubahan perilaku belanja masyarakat akibat pandemi COVID-19, yang membuat transaksi *online* semakin populer. Namun, pada tahun 2023, transaksi *e-commerce* mengalami sedikit penurunan sebesar 4,7% [7]. Penurunan ini bisa disebabkan oleh berbagai faktor, seperti pemulihan ekonomi pasca-pandemi, perubahan kebijakan impor, dan persaingan yang semakin ketat di industri ini. Meski demikian, pada tahun 2024, transaksi kembali meningkat menjadi Rp 487,01 triliun, menunjukkan bahwa sektor ini masih memiliki prospek cerah di masa depan. Transaksi *e-commerce* di Indonesia dapat dilihat pada Gambar 2.

Berdasarkan hasil penelusuran di media sosial Instagram, peneliti menemukan berbagai tanggapan dari pengguna terkait ulasan produk yang dijual di Shopee. Ulasan tersebut mencakup komentar positif maupun negatif, yang mencerminkan beragam persepsi konsumen terhadap kualitas produk, layanan penjual, maupun pengalaman berbelanja [8]. Analisis sentimen merupakan pendekatan yang digunakan untuk

mengungkap opini dan respons publik terhadap suatu objek, topik, atau fenomena tertentu. Dalam konteks penelitian ini, fokusnya adalah mengeksplorasi opini dan tanggapan konsumen terhadap ulasan produk Shopee yang tersebar di Instagram. Dengan memanfaatkan algoritma K-Nearest Neighbors (K-NN), penelitian ini diharapkan dapat memberikan gambaran yang lebih mendalam mengenai persepsi konsumen serta membantu pihak terkait dalam meningkatkan kualitas produk dan strategi pemasaran.



Gambar 2. Transaksi *E-Commerce* Indonesia 2019-2024

Pada penelitian terdahulu, fokus kajian diarahkan pada analisis sentimen terhadap berbagai platform *e-commerce* untuk memahami persepsi konsumen terhadap layanan maupun produk yang ditawarkan. Beberapa studi menggunakan metode *text mining* dan algoritma *machine learning* seperti Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM) dalam mengklasifikasikan ulasan konsumen menjadi kategori positif, negatif, atau netral [9]. Permasalahan yang umum diangkat adalah ketidaksesuaian antara ekspektasi konsumen dengan produk yang diterima, seperti kualitas barang, ketepatan pengiriman, dan respons layanan pelanggan.

Salah satu penelitian membandingkan performa algoritma K-NN dengan Naïve Bayes dalam klasifikasi ulasan produk pada platform *e-commerce*. Hasilnya menunjukkan bahwa K-NN memiliki akurasi rata-rata lebih tinggi, yaitu 96,9%, dibandingkan Naïve Bayes yang memperoleh akurasi sebesar 86,87% [10]. Penelitian lainnya memanfaatkan algoritma Decision Tree untuk mengelompokkan ulasan berdasarkan atribut tertentu, yang memudahkan analisis tren kepuasan konsumen terhadap produk. Berbeda dengan penelitian terdahulu tersebut, penelitian ini secara spesifik memfokuskan analisis pada ulasan produk Shopee yang dipublikasikan di Instagram [11]. Tujuannya adalah untuk memahami persepsi konsumen terhadap produk yang dijual di Shopee, dengan melihat tren sentimen yang berkembang di media sosial. Penelitian ini menggunakan algoritma K-NN sebagai metode klasifikasi, dengan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM) untuk pengolahan data. Evaluasi dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mengukur kinerja model [12].

Rujukan penelitian ini terletak pada kemampuannya memberikan wawasan yang komprehensif mengenai persepsi konsumen di media sosial, yang dapat digunakan oleh penjual dan pihak Shopee untuk meningkatkan kualitas produk, layanan, dan strategi pemasaran. Mengacu pada temuan penelitian sebelumnya, peningkatan akurasi dapat dicapai dengan memperbesar jumlah data ulasan yang dianalisis serta melakukan *parameter tuning* pada algoritma K-NN. Dengan pendekatan ini, diharapkan penelitian dapat memberikan hasil yang lebih akurat dan relevan, sekaligus menjadi referensi bagi penelitian selanjutnya dalam analisis sentimen pada sektor *e-commerce* [13].

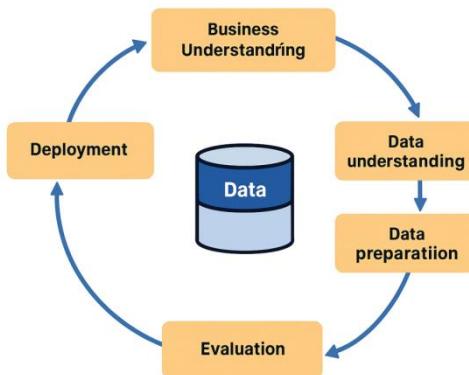
2. BAHAN DAN METODE

Dalam penelitian ini, penerapan metodologi dilakukan dengan menggunakan pendekatan CRISP-DM (lihat Gambar 3) untuk mengelola dan menganalisis data yang diperoleh dari Instagram terkait sentimen publik terhadap ulasan produk Shopee. Penelitian ini menggunakan metode kualitatif dengan dua sumber data utama, yaitu data primer dan data sekunder. Data primer diperoleh melalui observasi langsung, yaitu proses *crawling* data dari Instagram menggunakan teknik pengambilan data otomatis. Data sekunder dikumpulkan melalui studi literatur dari berbagai penelitian relevan yang tersedia di perpustakaan maupun sumber digital terpercaya [14].

Data primer dalam penelitian ini dikumpulkan melalui proses *crawling* ulasan atau komentar dari akun dan postingan di Instagram yang membahas produk Shopee. *Crawling* data adalah metode pengumpulan data secara otomatis dari media sosial yang berfokus pada opini dan sentimen pengguna. Pengumpulan data dilakukan dengan memanfaatkan *tools* dan *library* pemrograman untuk mengekstraksi komentar yang mengandung kata kunci atau tagar terkait produk Shopee. Proses ini melibatkan beberapa

langkah, termasuk autentikasi akses, pengambilan data, dan penyimpanan data mentah dalam format yang siap untuk dianalisis lebih lanjut. Data yang diambil mencakup teks komentar, tanggal posting, dan informasi akun pengguna yang relevan. Observasi ini memastikan bahwa data yang terkumpul merupakan data terbaru dan relevan dengan topik penelitian.

Data sekunder dalam penelitian ini diperoleh melalui studi literatur terhadap berbagai penelitian terdahulu yang terkait dengan analisis sentimen, algoritma K-NN, dan penerapan analisis sentimen pada media sosial, khususnya Instagram. Literatur yang digunakan berasal dari buku akademik, jurnal ilmiah, dan sumber terpercaya lainnya. Studi literatur ini bertujuan memberikan landasan teoritis yang kuat, memahami metodologi penelitian serupa, mengidentifikasi celah penelitian yang masih ada, serta memperkuat argumen dan hasil penelitian ini.



Gambar 3. Metode CRISP-DM

2.1. *Business Understanding*

Tahap awal dalam metodologi CRISP-DM adalah memahami konteks bisnis dan merumuskan tujuan analisis. Pada penelitian ini, tujuan utamanya adalah menganalisis sentimen publik terhadap penggunaan aplikasi Shopee berdasarkan ulasan yang diperoleh dari Instagram [15]. Analisis dilakukan untuk mengklasifikasikan opini pengguna ke dalam dua kategori, yaitu positif dan negatif. Hasil analisis diharapkan dapat memberikan gambaran mengenai persepsi masyarakat terhadap aplikasi Shopee, baik dari segi kelebihan maupun kekurangannya. Data yang dianalisis berupa komentar pada postingan Instagram yang membahas Shopee, sedangkan pemodelan klasifikasi sentimen dilakukan menggunakan algoritma K-NN.

2.2. *Data Understanding*

Data yang digunakan dalam penelitian ini berasal dari platform Instagram, yang dikumpulkan menggunakan teknik *crawling* dengan bantuan Google Extension sebagai alat pengambilan data. Proses pengumpulan data dilakukan dengan menyaring komentar yang mengandung kata kunci terkait penggunaan aplikasi Shopee. Dalam penelitian ini, sebanyak 1.100 komentar berhasil dikumpulkan melalui proses *crawling* tersebut. Setiap komentar yang dikumpulkan telah dipastikan relevansinya dengan topik penelitian, sehingga data yang diperoleh dapat memberikan informasi yang akurat dan sesuai dengan kebutuhan analisis. Setelah data terkumpul, langkah selanjutnya adalah mempelajari struktur, kualitas, dan distribusi data tersebut [16]. Analisis awal dilakukan untuk memahami karakteristik data, seperti panjang teks, variasi bahasa yang digunakan, serta konteks pembicaraan pengguna. Selain itu, dilakukan pula proses identifikasi untuk memastikan bahwa hanya data yang relevan dan bermakna yang akan digunakan dalam analisis sentimen. Tahap ini sangat penting untuk menjamin bahwa data yang digunakan memiliki kualitas yang baik dan dapat mendukung hasil penelitian yang valid.

2.3. *Data Preparation*

Pada tahapan ini, data yang dikumpulkan dari media sosial Instagram kemudian diproses lebih lanjut untuk diberikan label sentimen. Sentimen yang digunakan dalam penelitian ini terdiri dari dua kategori, yaitu positif dan negatif. Setelah proses pelabelan selesai, data tersebut memasuki tahap data *preprocessing*. Tahap ini meliputi pembersihan data dengan menghapus duplikasi, mention, tautan iklan, serta komentar yang tidak relevan dengan penelitian [17]. Selanjutnya dilakukan *transform case* untuk mengubah semua huruf menjadi huruf kecil agar format teks konsisten. Tahap pelabelan data bertujuan untuk mengkategorikan setiap komentar ke dalam sentimen positif atau negatif. Proses berikutnya adalah *tokenizing*, yaitu memecah kalimat menjadi kata-kata tunggal, kemudian *stopwords removal* untuk menghapus kata-kata umum yang tidak memberikan makna signifikan. Setelah itu, dilakukan *stemming* untuk mengubah kata yang berimbahan menjadi kata dasar. Terakhir, diterapkan *term weighting* menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk memberikan bobot pada setiap kata berdasarkan tingkat

kemunculannya di dalam dokumen, sehingga kata-kata yang memiliki pengaruh lebih besar terhadap sentimen dapat teridentifikasi dengan baik. Tahap ini sangat penting untuk memastikan data siap digunakan pada proses data preparation dan analisis menggunakan algoritma K-NN [18].

2.4. Modeling

Pada tahap ini, dilakukan proses pemodelan terhadap data komentar dari Instagram yang telah melalui tahapan pembersihan dan *preprocessing*. Data tersebut kemudian dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, dengan tujuan agar model dapat belajar dari sebagian data dan diuji pada data yang belum pernah digunakan sebelumnya. Proses ini terdiri dari dua langkah utama, yaitu pembagian data (*split data*) dan *cross-validation*. Dalam pembagian data, penelitian ini menggunakan rasio 80% untuk data *training* dan 20% untuk data *testing*. Selanjutnya, pada tahap *cross-validation*, data dibagi secara acak menjadi 10 bagian (*10-fold cross-validation*) untuk memastikan evaluasi model lebih menyeluruh dan meminimalkan risiko *overfitting* [19]. Setiap bagian digunakan secara bergantian sebagai data *testing* sementara sisanya sebagai data *training*. Untuk mengatasi ketidakseimbangan jumlah data pada masing-masing kelas sentimen, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Pemodelan dilakukan menggunakan algoritma K-NN untuk mengklasifikasikan komentar menjadi dua kategori, yaitu sentimen positif dan negatif. Algoritma K-NN dipilih karena kemampuannya dalam menangani data berbasis teks dan memberikan hasil yang optimal pada analisis sentimen dengan jumlah kelas yang terbatas. Berikut adalah persamaan dari algoritma K-NN:

$$d(x_i, x_j) = \sqrt{\sum (x_{ik} - x_{jk})^2} \quad (1)$$

2.5. Evaluation

Model yang telah dibangun kemudian dievaluasi untuk mengukur kemampuannya dalam mengklasifikasikan sentimen komentar dari Instagram. Proses evaluasi ini menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik tersebut digunakan untuk memberikan gambaran menyeluruh mengenai performa model, baik dari sisi ketepatan prediksi maupun keseimbangan antara data positif dan negatif [20]. Hasil evaluasi ini menjadi acuan untuk menilai sejauh mana model dapat memenuhi tujuan penelitian serta memberikan informasi yang relevan terkait pola sentimen pengguna terhadap aplikasi Shopee. Berikut adalah rumus yang digunakan pada perhitungan metrik evaluasi tersebut:

$$\text{accuracy} = (TP+TN)/(TP+FP+FN+TN) \quad (2)$$

$$\text{precision} = (TP)/(TP+FP) \quad (3)$$

$$\text{Recall} = (TP)/(TP+FN) \quad (4)$$

$$F1\ Score = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (5)$$

2.6. Deployment

Tahap *deployment* merupakan langkah terakhir dalam penelitian ini, yaitu menyajikan hasil dalam bentuk visualisasi data. Visualisasi ini bertujuan untuk menampilkan kumpulan kata yang paling sering muncul pada komentar Instagram, yang dikelompokkan berdasarkan sentimennya, positif maupun negatif. Dengan adanya visualisasi ini, proses analisis menjadi lebih efisien karena peneliti dapat memahami opini pengguna terhadap aplikasi Shopee tanpa harus membaca seluruh komentar satu-persatu.

3. HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil dan pembahasan mengenai kinerja K-NN yang diimplementasikan dengan metode CRISP-DM dalam penelitian ini. Setelah data komentar dari Instagram berhasil dikumpulkan, langkah berikutnya adalah melakukan analisis awal terhadap struktur, kualitas, serta distribusi data. Analisis ini mencakup pemeriksaan panjang teks komentar, variasi bahasa yang digunakan, serta tingkat interaksi pengguna. Selain itu, dilakukan proses pemilihan komentar yang relevan dengan topik penelitian untuk memastikan hanya data yang sesuai dan bermakna yang digunakan dalam analisis sentimen. Tahap ini penting untuk menjamin kualitas data sehingga hasil penelitian yang diperoleh dapat dipercaya dan valid.

3.1 Tahap Pre-processing

Sebelum digunakan pada proses klasifikasi, data komentar dari Instagram terlebih dahulu melalui tahap pre-processing. Tahap ini merupakan bagian penting dalam proses data mining karena berfungsi untuk mempersiapkan data mentah agar siap dianalisis. Pada awal tahap *pre-processing*, dilakukan proses pembersihan data dengan menghapus komentar yang bersifat duplikat, komentar spam, promosi, atau konten lain yang tidak berkaitan dengan penggunaan aplikasi Shopee. Setelah data dibersihkan dan difilter agar

hanya menyisakan komentar yang relevan, dilakukan proses *transform cases*, yaitu mengubah seluruh teks menjadi huruf kecil dengan tujuan menyamakan penulisan dan menghindari perbedaan makna akibat kapitalisasi huruf yang berbeda. Tahap berikutnya adalah *tokenizing*, yakni memecah kalimat komentar menjadi kata-kata tunggal untuk memudahkan proses analisis lebih lanjut. Proses ini kemudian dilanjutkan dengan *stemming*, yaitu mengubah kata yang memiliki imbuhan menjadi bentuk dasar atau kata akar agar setiap kata mempunyai bentuk yang seragam. Tahap terakhir dalam *pre-processing* adalah *stopword removal*, yaitu menghapus kata-kata umum yang sering muncul namun tidak memiliki pengaruh penting terhadap sentimen. Setelah melalui seluruh rangkaian *pre-processing* tersebut, diperoleh sebanyak 1.058 komentar yang telah siap digunakan dalam proses klasifikasi sentimen menggunakan algoritma K-NN.

Tabel 1. Tahap *Pre-processing*

Tahapan	Hasil
Data awal	Kenapa sih chat Shoppee suka gagal kirim gtu? Lemot banget @shopee_id Penangan Masalahnya saya juga ga jelas. Barang ga ada uang Tidak Kembali!!!
Cleaning	Kenapa sih chat Shoppee suka gagal kirim gtu Lemot banget shopee id Penangan Masalahnya saya juga ga jelas Barang ga ada uang Tidak Kembali
Transfrom Cases	kenapa sih chat shoppee suka gagal kirim gtu lemot banget shopee id penangan masalahnya saya juga ga jelas barang ga ada uang tidak kembali
Tokenizing	“kenapa”, “sih”, “chat”, “shoppe”, “suka”, “gagal”, “kirim”, “gtu”, “lemot”, “banget”, “shoppe”, “id” “penangan”, “masalahnya”, “saya”, “juga”, “ga”, “jelas”, “barang”, “ga”, “ada”, “uang”, “tidak”, “kembali”
Stemming	kenapa chat shoppee suka gagal kirim lemot banget shoppee id penangan masalah saya juga ga jelas barang ga ada uang tidak kembali
Stopwords	kenapa chat shoppee gagal kirim lemot shoppee penangan masalah ga jelas barang ga ada uang tidak kembali

Setelah proses *pre-processing* selesai (lihat Tabel 1), diperoleh 1.058 data komentar yang siap dianalisis. Langkah selanjutnya adalah melakukan pelabelan sentimen pada setiap komentar sesuai ekspresi yang terkandung di dalamnya (lihat Tabel 2). Pelabelan ini dilakukan secara manual terhadap 250 data yang akan digunakan sebagai data latih. Proses pelabelan dilakukan dengan bantuan *operator union*, dengan klasifikasi sentimen dibagi menjadi dua kategori, yaitu positif dan negatif. Hasil dari proses ini ditunjukkan pada tabel 2.

Tabel 2. Tahap Pelebelan

No	Label	Teks
1	Negatif	kenapa sih chat shoppee suka gagal kirim gtu lemot banget shopee id
2	Negatif	penangan masalahnya saya juga ga jelas barang ga ada uang tidak kembali

Tahap berikutnya adalah penerapan metode TF-IDF, yaitu teknik dalam pemrosesan bahasa alami yang digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah ulasan dibandingkan dengan keseluruhan kumpulan ulasan (*corpus*). Dalam penelitian ini, TF-IDF digunakan untuk mengidentifikasi kata-kata yang memiliki bobot signifikan dalam analisis sentimen terhadap ulasan produk Shopee yang diambil dari komentar pada Instagram. Hasil dari proses ini ditunjukkan pada Tabel 3.

Tabel 3. Tahap TF-IDF

No	Aplikasi	Shoppe	Banyak	Diskon	Keren
1	0,576	0,872	0,774	0,528	0,225
2	0,336	0,412	0,625	0,380	0,138
3	0,331	0,475	0,468	0,653	0,000
4	0,257	0,502	0,312	0,357	0,000
5	0,145	0,156	0,245	0,000	0,272
...
500	0,000	0,234	0,371	0,129	0,000

3.2 Pembuatan Model

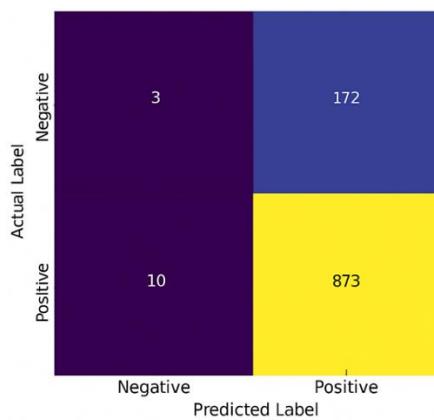
Pada tahap pemodelan, dilakukan proses pembentukan model klasifikasi yang mampu mengelompokkan sentimen ulasan produk Shopee di platform Instagram menggunakan algoritma K-NN. Berikut beberapa tahapan yang dilakukan:

1. Tahap pertama membuat model klasifikasi algoritma K-NN dan data latih yang nanti akan digunakan dalam proses sentimen analisis.

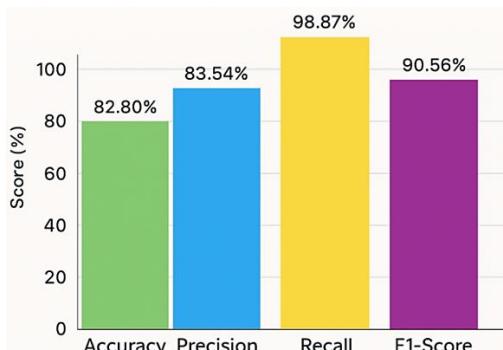
2. Tahap berikutnya adalah menggabungkan data latih (*training data*) dan data uji (*testing data*) untuk memastikan ketersediaan data yang siap digunakan pada pelatihan ulang model atau analisis lanjutan. Proses penggabungan ini dilakukan dengan menyatukan kedua set data menggunakan *operator union* yang sebelumnya telah dipisahkan, sehingga dapat memberikan gambaran yang lebih komprehensif mengenai kinerja model dan distribusi sentimen pada ulasan produk Shopee di Instagram.

3.3 Evaluasi Model

Untuk mengevaluasi kinerja model secara lebih mendalam, penelitian ini memanfaatkan *confusion matrix* yang menampilkan jumlah prediksi benar dan salah yang dihasilkan oleh model. Melalui *confusion matrix*, peneliti dapat menghitung metrik evaluasi seperti *precision*, *recall*, dan *F1-score*. Selain itu, untuk mengoptimalkan performa model, digunakan teknik SMOTE untuk mengatasi ketidakseimbangan kelas pada data ulasan. Pendekatan ini diterapkan untuk menunjukkan serta membuktikan hasil prediksi dari model K-NN dalam analisis sentimen ulasan produk Shopee di Instagram seperti gambar pada Gambar 4.



Gambar 4. Confusion Matrix



Gambar 5. Performance Matrix

Berdasarkan Gambar 5, *precision* menggambarkan tingkat ketepatan prediksi positif dibandingkan dengan jumlah seluruh prediksi positif yang dibuat oleh model. Berdasarkan hasil penelitian ini, presisi untuk ulasan bersentimen positif mencapai 83,54%, sedangkan presisi untuk ulasan bersentimen negatif adalah 23,08%. Hal ini menunjukkan bahwa dari semua ulasan yang diprediksi positif oleh model, 83,54% benar-benar positif, dan dari semua ulasan yang diprediksi negatif, 23,08% benar-benar negatif. *Recall* untuk sentimen negatif (*specificity*) adalah 1,71%, yang berarti model hanya mampu mengenali sebagian kecil ulasan yang benar-benar negatif. Sementara itu, *recall* untuk sentimen positif mencapai 98,87%, menunjukkan bahwa dari seluruh ulasan positif yang ada, model berhasil mengidentifikasi sebagian besar dengan benar. Nilai akurasi keseluruhan yang diperoleh menggunakan algoritma K-NN adalah 82,80%, sehingga dapat disimpulkan bahwa algoritma ini cukup baik dalam mengklasifikasikan sentimen ulasan produk di Instagram terkait Shopee.

3.4 Tahap Visualisasi

Tahap ini merupakan langkah akhir dalam proses analisis, yaitu pembuatan visualisasi hasil penelitian. Pada tahap ini, dibuat visualisasi data untuk menampilkan pola dan informasi penting yang diperoleh dari proses analisis sentimen ulasan produk Shopee di Instagram menggunakan algoritma K-NN. Visualisasi ini

bertujuan untuk mempermudah interpretasi hasil, sehingga temuan yang dihasilkan dapat dipahami dengan lebih jelas dan informatif.



Gambar 6. *World Cloud* Aplikasi Shoppe (a) negatif, (b) positif

Seperti pada Gambar 6, *Word Cloud* merupakan teknik visualisasi teks di mana kata-kata yang lebih sering muncul ditampilkan dengan ukuran lebih besar atau lebih menonjol, sedangkan kata-kata yang jarang muncul ditampilkan dengan ukuran lebih kecil. Teknik ini memudahkan dalam mengidentifikasi kata kunci serta tema yang dominan dalam kumpulan teks secara cepat dan intuitif. Pada penelitian ini, *Word Cloud* digunakan untuk menampilkan kata-kata yang paling sering muncul dalam ulasan produk Shopee di Instagram. Ukuran kata yang lebih besar menunjukkan tingkat kemunculan yang lebih tinggi, sehingga membantu mengungkap topik atau isu yang sering dibicarakan oleh pengguna.

4. KESIMPULAN

Berdasarkan hasil penelitian yang bertujuan untuk menganalisis sentimen ulasan produk Shopee di Instagram menggunakan algoritma K-NN, diperoleh akurasi sebesar 82,80%, dengan presisi sentimen positif mencapai 83,54% dan presisi sentimen negatif sebesar 23,08%. Hasil ini menunjukkan bahwa algoritma K-NN cukup efektif dalam mengklasifikasikan sentimen positif, namun memiliki keterbatasan dalam mengidentifikasi sentimen negatif secara tepat. Analisis sentimen mengungkap bahwa sentimen positif lebih dominan dibandingkan sentimen negatif. Dominasi sentimen positif ini menunjukkan bahwa mayoritas konsumen memberikan ulasan yang baik terhadap produk Shopee yang dipromosikan di Instagram, seperti kepuasan terhadap kualitas produk, kesesuaian dengan deskripsi, harga yang kompetitif, dan kecepatan pengiriman. Temuan ini menegaskan bahwa Shopee telah berhasil menciptakan pengalaman belanja yang memuaskan bagi sebagian besar penggunanya, meskipun masih terdapat ulasan negatif yang menyoroti permasalahan seperti produk yang tidak sesuai ekspektasi, keterlambatan pengiriman, atau respons penjual yang kurang memuaskan. Disarankan agar penelitian berikutnya menggunakan dataset yang lebih besar dan bervariasi untuk meningkatkan akurasi dan relevansi hasil. Selain itu, penggunaan algoritma pembanding dapat memberikan perspektif yang lebih komprehensif terhadap sentimen pengguna. Dari sisi teknis, pengembangan dashboard pemantauan sentimen secara *real-time* juga direkomendasikan. *Dashboard* ini memungkinkan pihak terkait untuk memantau opini publik secara langsung, mendeteksi isu sejak dini, dan merespons dengan cepat, sehingga citra dan kepercayaan konsumen terhadap platform dapat terjaga.

REFERENSI

- [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," *J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.
 - [2] S. Sahara, R. Agung Permana, U. Bina Sarana Informatika, and S. Antar Bangsa, "Metode Knn Pada Sentiment Analisis Review Produk Game Android," *Ijns.org Indones. J. Netw. Secur.*, vol. 11, no. 2, pp. 2302–5700, 2022, [Online]. Available: www.zdnet.com
 - [3] S. Pramuwidya, R. Y. Fa'rifah, and O. N. Pratiwi, "Klasifikasi Review Customer E-Commerce Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Bukalapak)," *e-Proceeding Eng.*, vol. 10, no. 3, pp. 3207–3212, 2023.
 - [4] D. Pajri, Y. Umaidah, and T. N. Padilah, "K-Nearest Neighbor Berbasis Particle Swarm Optimization untuk Analisis Sentimen Terhadap Tokopedia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–

- 253, 2020, doi: 10.28932/jutisi.v6i2.2658.
- [5] A. Ernawati, A. O. Sari, S. N. Sofyan, M. Iqbal, and R. F. W. Wijaya, "Implementasi Algoritma Naïve Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan," *Bull. Inf. Technol.*, vol. 4, no. 4, pp. 533–543, 2023, doi: 10.47065/bit.v4i4.1090.
- [6] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [7] M. R. R. Lillah, D. S. Maylawati, W. B. Zulfikar, W. Uriawan, and A. Wahana, "Implementasi Algoritma K-Nearest Neighbor (KNN) untuk analisis sentimen pengguna aplikasi Tokopedia," *Intellect Indones. J. Learn. Technological Innov.*, vol. 02, no. 02, pp. 171–184, 2023, [Online]. Available: https://digilib.uinsgd.ac.id/77056/%0Ahttps://digilib.uinsgd.ac.id/77056/6/4_BAB I.pdf
- [8] D. Sebastian, "Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace," *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 1, pp. 51–61, 2019, doi: 10.28932/jutisi.v5i1.1581.
- [9] M. F. El Firdaus, N. Nurfaizah, and S. Sarmini, "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1329, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [10] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [11] A. Rhamadanti, A. Rifa'i, F. Dikananda, and K. Anam, "Analisis Sentimen Pada Ulasan Access By Kereta Api Indonesia Dengan K-Nearest Neighbor," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3961.
- [12] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 98, 2023, doi: 10.53513/jsk.v6i1.7400.
- [13] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [14] Haris and D. Eka Ratnawati, "Analisis Sentimen berbasis Aspek terhadap Data Ulasan menggunakan Metode K-Nearest Neighbor (Studi Kasus: Aplikasi Olsera POS)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 3041–3046, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] J. A. Josen Limbong, I. Sembiring, K. Dwi Hartomo, U. Kristen Satya Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor Meth," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2021, doi: 10.25126/jtiik.202294960.
- [16] M. Saifurridho, U. Hayati, and Martanto, "Jurnal Informatika Terpadu Analisis Algoritma K-Nearest Neighbor Terhadap Sentimen Pengguna Aplikasi Shopee," *J. Inform. Terpadu*, vol. 10, no. 1, pp. 21–26, 2024, Available: <https://journal.nurulfikri.ac.id/index.php/JIT>
- [17] R. A. Permana and S. Sahara, "Algoritma K-Nearest Neighbor Pada Analisa Sentimen Review Produk Router," *Simkom*, vol. 8, no. 2, pp. 118–124, 2023, doi: 10.51717/simkom.v8i2.129.
- [18] A. D. Adhi Putra, "Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [19] M. Russell and M. Klassen, *Mining the social web: data mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and more.* 2019.
- [20] Ernianti Hasibuan and Elmo Allistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naïve Bayes Classifier," *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.