

Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor

Ikhsan Habib Kusuma^{1*)}, Nuri Cahyono²

^{1,2} Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom, Yogyakarta

^{1,2} Jln. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kab. Sleman, Daerah Istimewa Yogyakarta, 55281, Indonesia
email: ¹IkhsanKusuma@students.amikom.ac.id, ²NuriCahyono@amikom.ac.id

Abstract – *E-commerce's rapid growth has resulted in an increase in online transactions and shifts in consumer behavior. In Indonesia, the use of e-commerce has grown rapidly, with many online platforms emerging. Understanding public sentiment towards e-commerce in Indonesia is crucial for businesses to improve their services and maintain customer satisfaction. In this review, study propose a methodology for feeling investigation of popular assessment on the utilization of web-based business in Indonesia, utilizing directed learning calculations. The study involved collecting data from the website Google Play Store. The study performed data preprocessing, including removing stop words, tokenization, and stemming, before applying the K-Nearest Neighbor (K-NN) algorithm to classify sentiments into positive or negative. The evaluation was conducted using confusion matrix and classification report. The results showed that the proposed approach was effective in analyzing public sentiment towards e-commerce in Indonesia, with an accuracy rate of 82%. The study concluded that the proposed strategy could help businesses enhance their services and better satisfy customers' requirements and expectations.*

Keywords – *Sentiment Analysis, E-Commerce, Supervised Learning, Machine Learning, NLP, KNN.*

Abstrak – Perkembangan e-commerce yang pesat telah menyebabkan peningkatan transaksi online dan perubahan perilaku konsumen. Di Indonesia, penggunaan e-commerce tumbuh pesat dengan banyak platform online bermunculan. Memahami sentimen masyarakat terhadap e-commerce di Indonesia sangat penting bagi bisnis untuk meningkatkan layanan dan menjaga kepuasan pelanggan. Oleh karena itu, dalam penelitian ini peneliti mengusulkan sebuah pendekatan untuk melakukan analisis sentimen opini publik mengenai penggunaan salah satu e-commerce di Indonesia dengan menggunakan algoritma *K-Nearest Neighbor*. Pengumpulan data dilakukan dari website Google Play Store dengan tujuan untuk memperoleh pandangan dan pengalaman masyarakat terkait penggunaan salah satu e-commerce di Indonesia. Setelah data terkumpul, dilakukan proses preprocessing untuk membersihkan data, termasuk menghilangkan *stopwords*, tokenisasi, dan *stemming*. Setelah itu, algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk mengklasifikasikan sentimen menjadi positif atau negatif. Evaluasi dilakukan dengan menggunakan *confusion*

matrix dan *classification report* untuk menilai keakuratan algoritma. Hasil penelitian menunjukkan bahwa pendekatan yang diusulkan efektif dalam menganalisis sentimen masyarakat terhadap e-commerce di Indonesia, dengan tingkat akurasi 82%. Penelitian ini memiliki implikasi penting bagi bisnis e-commerce di Indonesia dalam meningkatkan layanan dan memenuhi kebutuhan serta harapan pelanggan secara lebih baik.

Kata Kunci – *Sentimen Analisis, E-Commerce, Supervised Learning, Machine Learning, NLP, KNN.*

I. PENDAHULUAN

Di Indonesia, penggunaan e-commerce berkembang pesat seiring dengan pergeseran perilaku konsumen, dengan semakin banyaknya orang yang melakukan pembelian secara online. Menurut laporan terbaru We Are Social, sekitar 178,9 juta orang Indonesia telah terlibat dalam belanja online dari awal tahun 2022 hingga akhir tahun 2023, menunjukkan pertumbuhan sebesar 12,8% dari tahun ke tahun. Hal ini juga didukung oleh perkiraan belanja online penduduk Indonesia yang mencapai US\$55,97 juta atau sekitar Rp 851 triliun pada tahun yang sama.

Namun, mempertahankan basis pengguna dan meningkatkan kualitas layanan platform e-commerce bergantung pada kepuasan pelanggan. Contoh-contoh platform e-commerce yang dapat dijadikan rujukan meliputi Shopee, Tokopedia, Lazada, Bukalapak, dan Blibli. Alasan pemilihan Shopee sebagai objek penelitian dalam konteks analisis sentimen masyarakat adalah karena popularitasnya yang mendominasi pasar e-commerce di Indonesia. Oleh karena itu, untuk memahami tingkat kepuasan dan kepercayaan pengguna terhadap platform ini, perlu dilakukan analisis sentimen publik terkait penggunaannya.

Dalam penelitian ini, peneliti mengusulkan pendekatan untuk melakukan analisis sentimen terhadap opini publik terhadap penggunaan Shopee dengan menggunakan algoritma *K-Nearest Neighbor*. Peneliti mengumpulkan data dari website google playstore dengan Teknik *web scraping* untuk kemudian dilakukan preprocessing data seperti penghapusan *stopwords*, *tokenization*, dan *stemming* sebelum menerapkan algoritma *K-Nearest Neighbor* (K-NN) dan *Natural Language Processing* (NLP) untuk mengklasifikasikan sentimen menjadi positif atau negatif. Evaluasi dilakukan dengan menggunakan *confusion matrix* dan *classification report*[1].

*) penulis korespondensi: Ikhsan Habib Kusuma
Email: ikhsankusuma@students.amikom.ac.id

II. PENELITIAN YANG TERKAIT

Beberapa penelitian yang terkait memiliki kaitan dengan penelitian ini yang pertama adalah Jurnal "*Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing*" yang ditulis oleh Muktafin E, Kusri K, dan Luthfi E menyediakan pandangan berharga dalam bidang analisis sentimen produk di lingkungan e-commerce. Penelitian ini bertujuan untuk mengotomatiskan klasifikasi sentimen terhadap ulasan produk, memungkinkan penjual dan pengguna untuk memahami pendapat pembeli mengenai produk yang ditawarkan.

Hasil penelitian menunjukkan bahwa penerapan algoritma *K-Nearest Neighbor* (KNN) dan *Natural Language Processing* (NLP) menghasilkan tingkat akurasi yang memuaskan, mencapai 76,92%. Ini meningkatkan presisi dan recall, masing-masing sekitar 80,00% dan 74,07%, dalam perbandingan dengan analisis tanpa NLP. Hasil ini menegaskan peran krusial NLP dalam meningkatkan kualitas analisis sentimen.

Dalam pengembangan penelitian ini, sebaiknya variasi produk yang dianalisis diperluas, yang dapat memberikan wawasan lebih luas tentang perilaku konsumen di platform Shopee. Terlebih lagi, jurnal ini memberikan sumbangan berharga untuk penelitian terkait, khususnya bagi mereka yang memerlukan informasi tentang bagaimana NLP dapat memengaruhi analisis sentimen ulasan produk e-commerce. Hal ini memungkinkan pemahaman yang lebih baik tentang cara pelanggan berinteraksi dalam e-commerce, yang dapat membantu dalam pengambilan keputusan bisnis online[1].

Selanjutnya, penelitian oleh Meishita Inelza Putri dan Iqbal Kharisudin (2022) tentang "*Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Support Vector Machine (SVM), Naive Bayes, dan Logistic Regression*". Hasilnya, data yang digunakan penulis berasal dari review aplikasi Tokopedia di Google Play, dengan total 3.125 review. Tujuan dari penelitian ini adalah menganalisis umpan balik pengguna menggunakan aplikasi Tokopedia dan menggabungkan hasil dari tiga metode klasifikasi yang berbeda.

Berdasarkan hasil penelitian, penggunaan metode SVM dalam klasifikasi sentimen pengguna pada aplikasi Tokopedia menunjukkan performa terbaik, dengan tingkat akurasi mencapai 90%. Selain itu, penulis menemukan bahwa 69% pengguna memiliki kesan positif terhadap aplikasi Tokopedia.

Meskipun demikian, penelitian Meishita Inelza Putri dan Iqbal Kharisudin (2022) tidak membahas secara khusus tentang analisis sentimen masyarakat terhadap penggunaan e-commerce secara umum[2].

Dalam penelitian sebelumnya yang dilakukan oleh Utami E (2023), berjudul "*Optimizing Sentiment Analysis of Product Reviews on Marketplace Using a Combination of Preprocessing Techniques, Word2Vec, and Convolutional Neural Network*." adalah pada analisis review produk di platform marketplace. Penelitian ini menggunakan teknik *Word2Vec*, *Convolutional Neural Network* (CNN), dan preprocessing untuk menganalisis 20.986 review produk dari 720 produk di marketplace Shopee.

Tingkat akurasi tertinggi sebesar 99,00%, persentase presisi tertinggi sebesar 98,96%, dan tingkat recall tertinggi sebesar 98,96%. Hal ini menggambarkan keberhasilan dalam

menerapkan teknologi NLP dan pengolahan data dalam analisis sentimen pada ulasan e-commerce. Oleh karena itu, penelitian yang terkait dengan judul jurnal "*Analisis Sentimen Masyarakat terhadap penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor*" perlu dilakukan untuk lebih memperdalam pemahaman tentang sentimen masyarakat terhadap penggunaan e-commerce di Indonesia[3].

III. METODE PENELITIAN

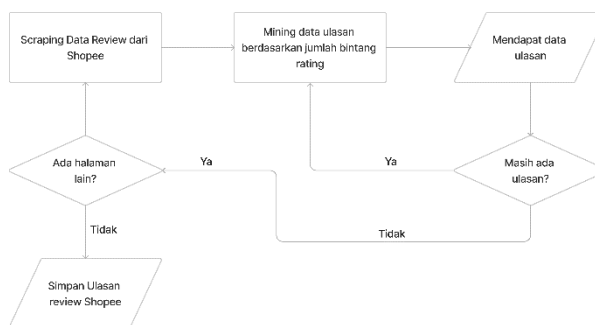
A. Web Scraping

Pada alur penelitian ini, dilakukan web scraping menggunakan WebHarvy pada platform Google Play Store dari google khususnya untuk aplikasi Shopee. Alur dimulai dengan memasukkan link ke halaman Shopee di Google Play Store yang ingin diambil data ulasan dan ratingnya.



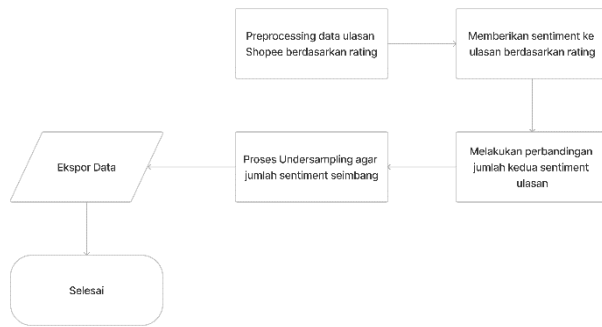
Gbr. 1 Proses memulai web scraping dengan menyalin link ke WebHarvy.

Setelah memasukkan link, WebHarvy akan melakukan scraping secara otomatis untuk mengumpulkan ulasan pengguna dan rating yang terdapat pada halaman tersebut.



Gbr. 2 Proses pengumpulan dataset dengan Teknik web scraping.

Data ulasan dan rating ini kemudian disimpan dalam format Excel untuk memudahkan pengolahan dan analisis selanjutnya[3].



Gbr. 3 Proses menyimpan dan preprocessing dataset hasil scraping.

Setelah proses scraping selesai, dilakukan tahap preprocessing data di Excel. Pada tahap ini, data ulasan dan rating akan diproses untuk mendapatkan sentimen positif dan negatif berdasarkan rating yang diberikan oleh pengguna. Misalnya, rating 4 atau 5 akan dianggap sebagai sentimen positif, sedangkan rating 1 atau 2 akan dianggap sebagai sentimen negatif.

Dalam tahap preprocessing ini, dilakukan pengelompokan rating menjadi dua kategori sentimen, yaitu positif dan negatif. Kemudian, data ulasan dan rating akan dibersihkan dari karakter-karakter yang tidak relevan dan dilakukan pembersihan data lainnya. Hal ini bertujuan untuk memastikan data yang digunakan dalam analisis sentimen sudah siap dan terstruktur dengan baik[4].

Setelah memberikan sentimen positif dan negatif berdasarkan rating di Excel pada dataset hasil web scraping dari Google Play Store Shopee, langkah selanjutnya adalah mengatasi ketidakseimbangan kelas. Undersampling digunakan untuk mengurangi jumlah sampel pada kelas mayoritas agar seimbang dengan jumlah sampel pada kelas minoritas[5].

B. Dataset

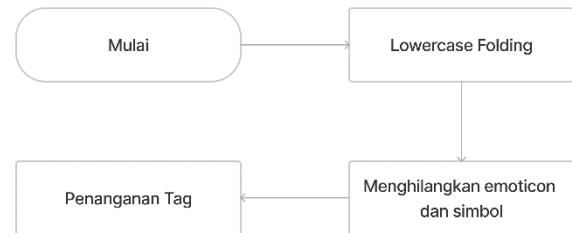
Penelitian ini menggunakan teknik web scraping untuk mengambil dataset dari website Google Playstore menggunakan software Webharvy khususnya aplikasi Shopee. Setelah dataset berhasil diambil, peneliti menggunakan library Pandas pada bahasa pemrograman Python untuk melakukan proses pembacaan dan penampilan dataset dalam bentuk tabel. Proses ini dilakukan untuk mendapatkan data yang dibutuhkan dan dapat digunakan dalam analisis sentimen masyarakat terhadap penggunaan e-commerce dengan menggunakan algoritma *K-Nearest Neighbor*[6].

Selanjutnya, dataset tersebut dilakukan proses undersampling untuk menjaga keseimbangan antara kategori positif dan negatif. Setelah dilakukan undersampling, dataset final terdiri dari 895 data, dengan 469 data yang diklasifikasikan sebagai sentimen positif dan 426 data yang diklasifikasikan sebagai sentimen negatif. Hal ini penting untuk memastikan keberimbangan dalam jumlah data pada setiap kategori sentimen, sehingga analisis dan evaluasi yang dilakukan dapat lebih objektif dan akurat.

C. Preprocessing Dataset

Dalam alur text processing untuk membersihkan data, terdapat beberapa tahapan yang dilakukan. Tahap pertama

adalah *lowercase folding*, di mana seluruh teks akan diubah menjadi huruf kecil agar tidak ada perbedaan dalam memproses kata yang sama dengan huruf besar dan huruf kecil.

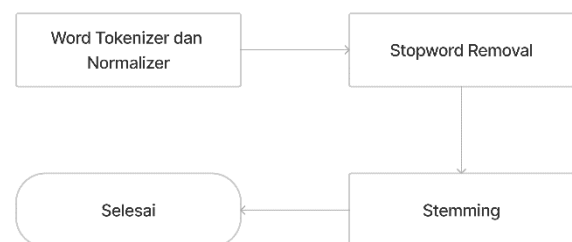


Gbr. 4 Tahap pra-pemrosesan NLP dengan cleaning simbol dataset.

Tahap berikutnya adalah menghilangkan tanda baca dan angka dari teks. Hal ini dikarenakan sering kali tidak memberikan kontribusi signifikan dalam analisis teks dan dapat mengganggu proses tokenisasi dan normalisasi kata. Dengan menghapus tanda baca dan angka, teks menjadi lebih bersih dan fokus pada kata-kata penting.

Setelah tahap penghapusan tanda baca dan angka dalam alur text processing, terdapat tahap penanganan tag dalam teks. Tag dalam teks dapat berupa tag HTML, XML, atau tag khusus lainnya yang terdapat dalam data yang diambil. Penanganan tag penting untuk menghilangkan elemen-elemen yang tidak diperlukan dan memastikan hanya teks yang relevan yang diproses[7].

Setelah itu, dilakukan tahap *word tokenization*, di mana teks akan dipisahkan menjadi token-token kata. Ini memungkinkan pemrosesan kata per kata untuk analisis selanjutnya. Berikutnya, dilakukan tahap *word normalization* untuk memperoleh bentuk dasar dari kata-kata tersebut. Misalnya, kata "berlari" akan dinormalisasi menjadi "lari" agar dapat dianggap sebagai bentuk yang sama.



Gbr. 5 Tahap pra-pemrosesan NLP dengan cleaning word dataset.

Selanjutnya, dilakukan tahap penghapusan *stopwords*, yaitu kata-kata umum yang sering muncul dalam teks namun tidak memberikan makna yang signifikan. *Stopwords* seperti "dan", "di", atau "untuk" akan dihapus agar fokus pada kata-kata kunci yang lebih informatif.

Terakhir, dilakukan *stemming* untuk mengubah kata-kata menjadi bentuk dasar atau kata akar. Kata-kata yang memiliki akar yang sama, seperti "membaca" dan "membacakan", dapat diubah menjadi "baca" melalui *stemming*.

Dengan mengikuti alur text processing yang mencakup *lowercase folding*, penghapusan tanda baca dan angka, *word tokenization*, *word normalization*, penghapusan *stopwords*, dan *stemming*, data teks menjadi lebih terstruktur dan siap untuk dilakukan analisis lebih lanjut seperti analisis sentimen. Proses ini memastikan bahwa data penelitian telah melalui prosedur pembersihan yang sesuai untuk menghasilkan data yang lebih relevan dan akurat[8].

Pada tahap *feature extraction* dilakukan ekstraksi fitur (feature) dari data yang telah diproses pada tahap sebelumnya. Beberapa fitur yang dapat diekstraksi meliputi kata-kata (*term frequency*), bobot kata TF-IDF (*term weighting*), dan kategori kata (*part of speech tagging*). Pada tahap pelatihan model, model machine learning dilatih menggunakan data training.

Pada tahap evaluasi, model yang telah dilatih dinilai menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan confusion matrix. Evaluasi ini membantu dalam mengukur kinerja dan efektivitas model yang telah dilatih. Terakhir pada tahap deployment, model yang telah terlatih dapat diintegrasikan ke dalam aplikasi atau sistem yang digunakan untuk memantau sentimen pelanggan terhadap produk atau layanan[9].

D. Algoritma

NLP (*Natural Language Processing*) adalah cabang kecerdasan buatan yang berfokus pada pemahaman, pengolahan, dan generasi bahasa manusia oleh komputer. Pada tahap ini, teks diubah menjadi *lowercase*, tanda baca dan angka dihilangkan, dan kata-kata dipisahkan menjadi token. Hal ini mempersiapkan data teks agar dapat diolah lebih lanjut[10].

Teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) memperkirakan seberapa penting sebuah kata dalam dokumen atau koleksi dokumen. Rumus kalkulasi TF-IDF menggabungkan frekuensi kata dalam dokumen dengan inversi frekuensi kata dalam keseluruhan dokumen, sehingga memberikan bobot yang lebih tinggi pada kata-kata yang jarang muncul tapi memiliki makna penting[11].

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

Keterangan:

$tf_{i,j}$ = banyaknya kata-i pada dokumen ke-j

N = total dokumen

df_i = banyaknya dokumen yang mengandung kata ke-i

Selanjutnya, terdapat metode *Cosine Similarity* yang digunakan untuk mengukur sejauh mana dua teks mirip satu sama lain. Metode ini menghitung kemiripan antara dua vektor representasi teks dengan menggunakan rumus kalkulasi *cosine similarity*. Semakin tinggi nilai *cosine similarity*, semakin mirip kedua teks tersebut.

K-Nearest Neighbors (KNN) digunakan sebagai algoritma *supervised learning* untuk melakukan klasifikasi sentimen positif dan negatif dari komentar masyarakat terhadap penggunaan e-commerce. KNN adalah algoritma klasifikasi yang sederhana namun efektif, yang berdasarkan pada konsep bahwa data dengan atribut yang mirip cenderung memiliki label kelas yang sama[12].

Proses KNN dimulai dengan memilih sebuah titik atau observasi dari data training yang telah dilabeli. Kemudian, KNN mencari K tetangga terdekat dari titik yang dipilih, di mana K adalah nilai yang ditentukan sebelumnya oleh pengguna. Setelah itu, algoritma KNN akan menghitung jarak antara titik yang dipilih dengan K tetangga terdekat menggunakan metrik jarak tertentu seperti *Euclidean distance* atau *Manhattan distance*. Akhirnya, label kelas yang paling sering muncul dari K tetangga terdekat akan diambil sebagai label kelas untuk titik yang dipilih.

Setelah dilakukan preprocessing data, KNN dilatih dengan data training untuk menentukan nilai K dan membangun model klasifikasi. Setelah model dibangun, data testing digunakan untuk menguji akurasi model. Setelah itu, model dapat digunakan untuk mengkategorikan data yang belum dilihat [13].

Untuk mengevaluasi performa model klasifikasi, terdapat *confusion matrix* yang membandingkan hasil klasifikasi aktual dengan prediksi yang diberikan oleh model. Ada empat nilai penting dalam *matrix confusion*: Nilai Positif Benar (TP), Nilai Negatif Benar (TN), Nilai Negatif Benar (FP), dan Nilai Negatif Benar (FN)[14].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{FP+TP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{FN+TP} \times 100\% \quad (4)$$

E. Evaluasi

Confusion matrix digunakan untuk mengevaluasi kinerja dari algoritma klasifikasi yang digunakan untuk melakukan analisis sentimen. *Matrix confusion* terdiri dari empat bagian: Nilai Positif Benar (TP) adalah jumlah prediksi yang benar dan nilai aktualnya juga benar; Nilai Negatif Benar (TN) adalah jumlah prediksi yang benar dan nilai aktualnya juga benar; Nilai Negatif Benar (FN) adalah jumlah prediksi yang salah dan nilai aktualnya juga benar; dan Nilai Negatif Benar (FN) adalah jumlah prediksi yang salah dan nilai aktualnya juga benar[15].

TABLE I
CONFUSION MATRIX

Confusion Matrix	Hasil Klasifikasi
True Positif	75
False Negatif	19
False Positif	12
True Negatif	73

Accuracy, *Precision*, *Recall*, *F1-Score*, dan *Support* adalah metrik evaluasi penting dalam tugas klasifikasi. *Precision* mengukur persentase instance positif yang benar diklasifikasikan oleh model, *recall* mengukur persentase instance positif yang berhasil ditemukan, dan *F1-Score* adalah gabungan dari *precision* dan *recall*. *Support* adalah jumlah kemunculan aktual dari setiap kelas dalam dataset. Metrik-metrik ini memberikan gambaran tentang performa model dalam mengklasifikasikan data dengan akurat dan seimbang.

TABLE II
METRIK EVALUASI

	Precision	Recall	F1-Score	Support
Negatif	0.79	0.86	0.82	85
Positif	0.86	0.80	0.83	94
Accuracy			0.83	179
Macro Avg	0.83	0.83	0.83	179
Weighted Avg	0.83	0.83	0.83	179

Accuracy: 0.8268156424581006

Precision: 0.8620689655172413

Recall: 0.7978723404255319

F1 Score: 0.8287292817679558

Error Rate: 0.17318435754189943

Metode *Cross Validation Score* adalah untuk memperkirakan bagaimana model akan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Dalam *Cross Validation*, dataset yang ada akan dibagi menjadi beberapa subset yang saling bertumpuk dan bergeser, lalu model akan dilatih pada subset yang lebih besar dan diuji pada subset yang lebih kecil. Pada umumnya, teknik yang digunakan adalah *K-Fold Cross Validation*, di mana dataset dibagi menjadi k subset yang sama ukurannya.

Dataset harus dibagi menjadi k bagian yang sama ukurannya untuk menerapkan validasi *cross-fold K-fold*. Setelah itu, model akan dilatih pada bagian $k-1$ dan diuji pada bagian yang tersisa. Proses ini akan diulang sebanyak k kali, dengan setiap bagian berfungsi sebagai bagian uji tepat sekali. Hasil akhir pengujian cross validation adalah akurasi rata-rata dari setiap iterasi pengujian *cross-fold K-Fold*. Teknik ini memungkinkan pengujian model dengan lebih akurat, membuatnya lebih dapat diandalkan saat membuat keputusan.

IV. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Data yang digunakan dalam penelitian ini adalah dataset review pengguna aplikasi Shopee yang diambil dari Google Play Store menggunakan teknik *web scraping* dengan software WebHarvy. Data ini terdiri dari 895 baris dengan kolom Review dan Sentimen.

B. Preprocessing Data

Proses pembersihan data dilakukan pada tahap preprocessing data dengan mengubah *lowercase folding*, menghilangkan tanda baca, dan menghilangkan karakter yang tidak diperlukan. Kemudian, interaksi tokenisasi selesai, khususnya memecah kalimat menjadi token atau kata. Selain itu, *stopword removal* dan *stemming* digunakan untuk membuat kata-kata dasar yang relevan.

C. Model Sentiment Analysis

Model Sentiment Analysis menggunakan algoritma NLP dan KNN dalam penelitian ini. Proses training dibagi menjadi dua bagian, yaitu training set dan testing set dengan proporsi

80:20. Selanjutnya, dilakukan ekstraksi elemen dengan teknik TF-IDF untuk menghitung bobot setiap kata dalam laporan. Pada tahap training, algoritma KNN digunakan untuk mengklasifikasikan setiap dokumen ke dalam kelas sentimen positif atau negatif. Dari hasil penelitian, disimpulkan bahwa model ini dapat melakukan analisis sentimen pada dataset review pengguna aplikasi Shopee dengan akurasi sebesar 82% dan nilai *cross validation score* sebesar 80%. Hal ini menandakan bahwa kombinasi algoritma NLP dan KNN dapat menghasilkan hasil analisis sentimen yang baik.

D. Evaluasi Model

Untuk mengevaluasi model yang telah dibuat, digunakan metode *confusion matrix* dan *cross validation score*. *Confusion matrix* digunakan untuk menunjukkan jumlah hasil klasifikasi yang benar dan salah dari model. Sedangkan, *cross validation score* digunakan untuk mengukur seberapa baik model yang dibuat dalam memprediksi data yang belum dilihat sebelumnya.

E. Visualisasi

Pada penelitian ini, dilakukan visualisasi data menggunakan wordcloud untuk menganalisis frekuensi tentang kata-kata yang paling dominan dalam ulasan pengguna, sehingga memudahkan pemahaman mengenai preferensi dan pengalaman pengguna terkait dengan penggunaan e-commerce. Visualisasi *wordcloud* ini membantu dalam mengidentifikasi kata-kata kunci yang paling signifikan dan memperoleh wawasan yang berguna dalam analisis sentimen pada dataset tersebut.



Gambar 1. Wordcloud frekuensi kata yang sering muncul

V. KESIMPULAN

Dari hasil penelitian, dapat disimpulkan bahwa *model Sentiment Analysis* yang dikembangkan menggunakan algoritma NLP dan KNN menunjukkan kinerja yang memuaskan. Akurasi model mencapai 82%, dan nilai dari *cross-validation score* sebesar 80%. Evaluasi lainnya juga memberikan hasil yang positif, menandakan bahwa model ini efektif dalam melakukan analisis sentimen pada dataset ulasan pengguna aplikasi Shopee. Hasil ini memberikan keyakinan bahwa kombinasi algoritma NLP dan KNN dapat diandalkan untuk tugas analisis sentimen pada dataset yang relevan.

Namun, terdapat beberapa faktor yang dapat mempengaruhi akurasi model, seperti kualitas data, pemilihan fitur, dan parameter pada algoritma. Oleh karena itu, untuk

meningkatkan akurasi model, penelitian selanjutnya dapat dilakukan dengan meningkatkan aspek-aspek tersebut.

Selain itu, eksplorasi lebih lanjut dapat diselesaikan dengan menggunakan teknik yang lebih rumit, seperti *deep learning*, untuk meningkatkan akurasi dan performa dari model. Metode ini dapat digunakan untuk melakukan ekstraksi fitur yang lebih kompleks dan dapat memperbaiki hasil analisis sentimen yang dihasilkan.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan rasa terima kasih yang tulus kepada semua orang yang membantu, terutama dari universitas dan pembimbing. Kontribusi mereka sangat dihargai dan berperan penting dalam kesuksesan penulisan ini. Semua kontribusi ini menjadi pilar penting dalam keberhasilan penelitian ini, dan penulis sangat menghargai semua bantuan yang diberikan.

DAFTAR PUSTAKA

- [1] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal Eksplorasi Informatika*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplorasi.v10i1.390.
- [2] M. I. Putri and I. Kharisudin, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, pp. 759–766, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [3] E. Utami, "OPTIMIZING SENTIMENT ANALYSIS OF PRODUCT REVIEWS ON MARKETPLACE USING A COMBINATION OF PREPROCESSING TECHNIQUES, WORD2VEC, AND CONVOLUTIONAL NEURAL NETWORK OPTIMISASI ANALISIS SENTIMEN ULASAN PRODUK PADA MARKETPLACE DENGAN KOMBINASI TEKNIK PREPROCESSING, WORD2VEC, DAN CONVOLUTIONAL NEURAL NETWORK," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 1, pp. 101–107, 2023, doi: 10.20884/1.jutif.2023.4.1.815.
- [4] I. M. Wiryana, S. Harmanto, A. Fauzi, I. B. Qisthi, and Z. N. Utami, "Store product classification using convolutional neural network," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1439–1447, Sep. 2023, doi: 10.11591/ijece.v12i3.pp1439-1447.
- [5] A. Sri Widagdo, B. W. Soedijono A, and A. Nasiri, "Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes," 2020.
- [6] S. Hardani and D. Ajeng Kristiyanti, "Systematic Literature Review: Analisa Sentimen Penerimaan Masyarakat Terhadap Jenis Vaksin Covid-19 di Dunia." [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/issue/archive>
- [7] J. Mantik, Y. R. Saputri, and H. Februariyanti, "SENTIMENT ANALYSIS ON SHOPEE E-COMMERCE USING THE NAÏVE BAYES CLASSIFIER ALGORITHM," 2022.
- [8] B. Z. Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [9] M. Demircan, A. Seller, F. Abut, and M. F. Akay, "Developing Turkish sentiment analysis models using machine learning and e-commerce data," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 202–207, Jun. 2021, doi: 10.1016/j.ijcce.2021.11.003.
- [10] N. Khotimah, M. Yamin Darsyah, and I. M. Nur, "ANALISIS SENTIMEN TERHADAP REVIEW E-COMMERCE DENGAN METODE STOCHASTIC GRADIENT DESCENT." [Online]. Available: <http://repository.unimus.ac.id>
- [11] F. D. Ananda and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 407–416, May 2021, doi: 10.30812/matrik.v20i2.1130.
- [12] A. Bisri and M. Man, "INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION journal homepage: www.joiv.org/index.php/joiv INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION Machine Learning Algorithms Based on Sampling Techniques for Raisin Grains Classification," 2023. [Online]. Available: www.joiv.org/index.php/joiv
- [13] Junadhi, Agustin, M. Rifqi, and M. K. Anam, "Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 216–225, Dec. 2022, doi: 10.23887/janapati.v11i3.51531.
- [14] Y. Dani and M. A. Ginting, "INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION journal homepage: www.joiv.org/index.php/joiv INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION Classification of Predicting Customer Ad Clicks Using Logistic Regression and k-Nearest Neighbors," 2023. [Online]. Available: www.joiv.org/index.php/joiv
- [15] "Sentiment Analysis of OYO App Reviews Using the Support Vector Machine Algorithm Analisis Sentimen terhadap Ulasan Aplikasi OYO menggunakan Algoritma Support Vector Machine Zaenal, Ika Ratna Indra Astutik."