

Implementation of Aspect-Based Sentiment Analysis for Outdoor Gear Product Reviews on the Tokopedia E-Commerce Platform (Case Study: Eiger Adventure)

Dhanif Daffa Alfaridzi¹, Tacbir Hendro Pudjiantoro², Irma Santikarama³

^{1,2,3}Program Studi Sistem Informasi, Universitas Jenderal Achmad Yani, Indonesia

Email: dda.21@si.unjani.ac.id; tacbir.hendro@lecture.unjani.ac.id; irma.santikarama@lecture.unjani.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengungkap persepsi pelanggan terhadap produk *Eiger Adventure* di platform *Tokopedia* melalui pendekatan *Aspect-Based Sentiment Analysis* (ABSA). Data berupa 3.670 ulasan pelanggan dikumpulkan menggunakan teknik *web scraping* dan diproses melalui tahapan *preprocessing*, ekstraksi aspek, pelabelan sentimen, serta klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Proses vektorisasi teks dilakukan dengan metode *TF-IDF*, sedangkan metrik evaluasi meliputi akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa aspek yang paling dominan adalah kualitas, model, dan pengiriman. Sentimen positif paling banyak ditemukan pada aspek kualitas dan model, sementara aspek pengiriman dan ukuran menjadi sumber keluhan utama. Model klasifikasi mencapai akurasi sebesar 61,6% dan bekerja paling baik dalam mendeteksi sentimen positif. Sebagai implementasi, dikembangkan sistem berbasis *web* yang mampu melakukan analisis ABSA secara terintegrasi dan interaktif. Penelitian ini menyimpulkan bahwa meskipun produk *Eiger* dinilai unggul secara kualitas, upaya peningkatan layanan non-produk seperti pengiriman dan informasi ukuran sangat dibutuhkan untuk meningkatkan kepuasan pelanggan secara menyeluruh.

Keyword: Analisis Sentimen Berbasis Aspek; Tokopedia; KNN

ABSTRACT

This study aims to uncover customer perceptions of Eiger Adventure products on the Tokopedia platform through an Aspect-Based Sentiment Analysis (ABSA) approach. Data in the form of 3,670 customer reviews were collected using web scraping techniques and processed through preprocessing, aspect extraction, sentiment labeling, and classification using the K-Nearest Neighbor (KNN) algorithm. The text vectorization process was carried out using the TF-IDF method, while evaluation metrics included accuracy, precision, recall, and F1-score. The results showed that the most dominant aspects were quality, model, and delivery. Positive sentiment was mostly found in the quality and model aspects, while delivery and size aspects were the main sources of complaints. The classification model achieved an accuracy of 61.6% and performed best in detecting positive sentiment. As an implementation, a web-based system was developed that is capable of conducting ABSA analysis in an integrated and interactive manner. This study concluded that although Eiger products are considered superior in quality, efforts to improve non-product services such as delivery and size information are urgently needed to improve overall customer satisfaction.

Keyword: Aspect Based Sentiment Analysis; Tokopedia; KNN

Corresponding Author:

Dhanif Daffa Alfaridzi,
Universitas Jenderal Achmad Yani,
Jl. Terusan Jenderal Sudirman, Cimahi, Jawa Barat, Kota Cimahi, Jawa Barat
40525, Indonesia
Email: dda.21@si.unjani.ac.id



1. INTRODUCTION

Perdagangan berbasis elektronik merupakan contoh nyata dari perubahan cara manusia berinteraksi dan bertransaksi. Perubahan ini terjadi berkat pesatnya pertumbuhan penggunaan internet dalam kehidupan

sehari-hari, baik secara global maupun di Indonesia. Berdasarkan laporan *Digital 2025: Indonesia* (Kemp, 2025), sebanyak 74,6% penduduk Indonesia atau sekitar 212 dari 250 juta orang telah terhubung ke internet pada Januari 2025. Angka ini menunjukkan pertumbuhan sebesar 8,7% dibandingkan tahun-tahun sebelumnya. Penelitian oleh Dewi dan Lusikooy (2024) juga melaporkan bahwa nilai transaksi *e-commerce* di Indonesia, atau *Gross Merchandise Value (GMV)*, mengalami pertumbuhan rata-rata sebesar 20% per tahun selama periode 2023 hingga 2024. Bukti ini mencerminkan lonjakan aktivitas belanja digital yang sangat signifikan. Selain itu, Afonso et al. (2023) menunjukkan bahwa pandemi COVID-19 telah mendorong percepatan penggunaan platform digital hingga 30% lebih tinggi selama 2020–2022 dibandingkan masa sebelum pandemi. Angka ini menunjukkan bahwa hampir tiga perempat penduduk Indonesia telah memiliki akses terhadap internet, yang berjalan selaras dengan pertumbuhan ekosistem digital dan aktivitas belanja daring.

Penelitian oleh Shankar et al. (2021) mengidentifikasi beberapa faktor kunci dalam meningkatnya penggunaan *e-commerce*, yaitu kemudahan akses, keterbukaan harga, dan personalisasi konsumen. Faktor-faktor tersebut juga meningkatkan keterlibatan konsumen, yang mengakibatkan bertambah banyaknya ulasan terhadap produk yang telah dibeli. Namun, di sisi lain, terdapat ketidaksesuaian yang cukup sering terjadi antara ulasan produk dan tingkat kepuasan pelanggan yang sebenarnya. Penelitian menunjukkan bahwa komentar atau ulasan serta *rating* penilaian memang memengaruhi persepsi dan keputusan pembelian, sekaligus berfungsi sebagai cerminan citra merek tersebut (Rachmiani et al., 2023). Akan tetapi, hasil penelitian juga menunjukkan bahwa perbedaan antara ulasan dan kepuasan sebenarnya bersifat asimetris dan dipengaruhi oleh berbagai faktor lain, seperti waktu penulisan ulasan dan pengalaman individu (Xu et al., 2024).

Banyaknya data ini menjadikan ulasan pelanggan sebagai bagian penting dalam ekosistem belanja daring (Natasya et al., 2024). Setiap hari, jutaan ulasan dihasilkan di berbagai platform *e-commerce*, menciptakan volume data teks yang sangat besar. Data teks yang besar ini merupakan aset yang berharga bagi perusahaan karena dapat memberikan wawasan penting mengenai preferensi dan kepuasan pelanggan terhadap produk (Camilleri & Filieri, 2023; Changchit & Klaus, 2020). Namun, hal ini juga menimbulkan tantangan besar apabila dianalisis secara manual (Amalia & Winarko, 2021). Proses membaca dan menginterpretasi setiap ulasan satu per satu menjadi tidak efisien, memakan waktu, rawan kesalahan, serta rentan terhadap subjektivitas.

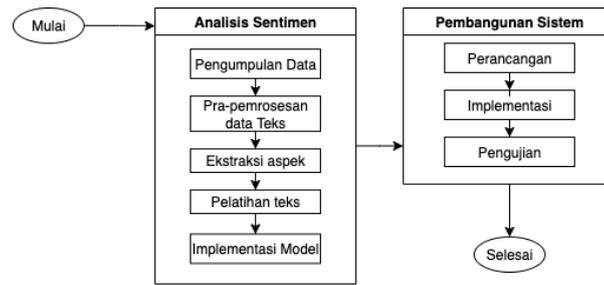
Selain memakan waktu, banyaknya data juga menyebabkan kesulitan dalam memahami opini karena bentuk datanya sangat beragam dan tidak terstruktur. Maka dari itu, penelitian ini mengusulkan penggunaan analisis sentimen untuk memperoleh pemahaman yang lebih baik mengenai pasar dan pelanggan (Syarif Imron et al., 2023). Namun, penelitian juga menunjukkan bahwa ulasan yang sangat argumentatif atau detail dari satu pelanggan bisa lebih berpengaruh dibandingkan skor rata-rata, meskipun tidak selalu mewakili tingkat kepuasan secara keseluruhan (Ziegele & Weber, 2015). Fakta ini menunjukkan bahwa selain kuantitas data, kualitas informasi dalam ulasan juga sangat penting.

Oleh karena itu, diperlukan pendekatan analisis yang lebih mendalam untuk menangkap makna di balik opini pelanggan secara lebih akurat. Hal ini karena analisis sentimen konvensional umumnya hanya bekerja di tingkat dokumen dan sering kali gagal menangkap nuansa yang kompleks, terutama ketika satu ulasan memuat penilaian yang berbeda terhadap berbagai aspek produk (Chris et al., 2024). Hasil penelitian sebelumnya menunjukkan bahwa *Aspect-Based Sentiment Analysis (ABSA)* secara konsisten mengungguli analisis sentimen konvensional dalam lima *benchmark*, dengan peningkatan presisi dan *recall* yang signifikan (Shukla & Dwivedi, 2025). Analisis sentimen konvensional yang hanya menghasilkan satu kesimpulan “positif” atau “negatif” untuk keseluruhan ulasan akan kehilangan informasi penting ini (Jazuli et al., 2023).

2. RESEARCH METHOD

A. Desain Penelitian

Desain penelitian dirancang sebagai rangkaian yang menjelaskan alur pengerjaan pada penelitian ini. Proses dimulai dengan *Pengumpulan Data* ulasan dari *e-commerce*. Data mentah tersebut kemudian akan melalui tahap *Memproses Data Teks* yang mencakup serangkaian langkah *pre-processing*. Setelah data bersih, akan dilakukan *Ekstraksi Aspek* untuk mengidentifikasi fitur produk yang sering dibicarakan, yang kemudian dilanjutkan dengan *Analisis Sentimen* untuk menentukan opini pada setiap ulasan dengan aspek tertentu. Kinerja dari keseluruhan alur kerja ini akan diukur pada tahap *Evaluasi model*, setelah itu model-nya diimplementasi pada keseluruhan data. Terakhir, akan dibuatkan sebuah sistem sederhana untuk memvisualisasikan proses penelitian yang sudah dilakukan.



Gambar 1. Desain Penelitian

B. Dataset

Data yang digunakan dalam penelitian ini merupakan data primer, ulasan pelanggan yang dikumpulkan dari toko resmi (*Official Store*) Eiger Adventure di platform *e-commerce* Tokopedia. Proses pengumpulan data dilakukan dengan teknik *web scraping* selama periode 1 Januari hingga 20 Juni 2025, memanfaatkan modul *requests* untuk mengirim permintaan *HTTP* dan *BeautifulSoup* untuk *parsing* atau membaca isi di halaman yang sedang dibuka.



Gambar 2. Sumber Data

Field yang bisa diambil adalah `scraped_at`, `review_date`, `product_name`, `product_variant`, `reviewer_name`, `rating`, dan `review_text`. Namun disini hanya mempertahankan field penting yaitu `review_text`.

Tabel 1. Data Mentah

review_text
pas di tangan untuk ukuran XL, dan tentunya nyaman utk motoran karena jari bisa untuk tap layar HP. Makasi 🙏🙏🙏 Makasih min sudah sampai. Pengiriman cepat, 2 hari sudah sampai padahal beda pulau. Dapat harga diskon sekitar 41%. Sepatu nya bagus, ga nyesal beli. Moga awet kualitas sudah TDK di ragukan lagi 🙏 produk original penjual ramah n gercep simple dan keren modelnya. makasih seller dan tokped juga kurirnya yg gercep 🙏🙏 lingkar kepala 58 cocok dan nyaman pake size M produk berkualitas penjual responsif pengiriman cukup lama

C. Pre-Processing

Data yang sudah dikumpulkan kemudian dilakukan *pre-processing* untuk menghindari data yang belum siap untuk diolah, seperti terdapat gangguan (*noise*) dan data yang tidak konsisten. Tahapannya yaitu *cleaning*, menghapus simbol, bilangan angka, dan tanda baca tersebut karena hanya teks alfabet yang akan digunakan. Lalu, kata-kata yang tidak bermakna atau tidak memiliki arti seperti kata *saya*, *dan*, *atau*, akan dilakukan penghapusan pada tahap *stopword-removal*. *Tokenization* adalah tahap memecah kalimat menjadi bagian-bagian terkecilnya yang dinamakan *token*. Terakhir, *stemming* adalah tahap mengubah sebuah kata ke dalam bentuk kata dasarnya dengan menghapus imbuhan di depan maupun imbuhan di belakang kata.

Setelah melewati semua tahap di atas, data menjadi bersih dan siap digunakan pada tahap selanjutnya.

Tabel 2. Data Siap Pakai

cleaned_review
['pas','tangan','ukur','xl','tentu','nyaman','motor','karena','jari','bisa','tap','layar','hp','makasih'] ['makasih','min','sudah','sampai','kirim','cepat','hari','sudah','sampai','beda','pulau','dapat','harga','diskon','sekitar','sepatu','bagus','nyesal','beli','moga','awet'] ['kualitas','sudah','ragu','lagi','produk','original','penjual','ramah','gercep'] ['simple','keren','model','makasih','seller','tokped','kurir'] ['lingkar','kepala','cocok','nyaman','pakai','produk','kualitas','penjual','responsif','kirim','cukup','lama']

D. Ekstraksi Aspek

Ekstraksi aspek dilakukan untuk mengidentifikasi elemen produk yang dikomentari oleh pengguna pada teks ulasan. Setiap aspek adalah kata atau frasa yang menunjukkan fitur spesifik yang akan dianalisis nilai sentimennya. Proses ini memisahkan opini pengguna berdasarkan atribut produk, sehingga menghasilkan wawasan yang lebih terperinci untuk dianalisis lebih lanjut.

Langkah pertama di proses ini adalah *part-of-speech tagging*, yaitu dengan cara mengambil frasa atau kata yang paling sering muncul dan dianggap penting oleh konsumen. Data ulasan diperiksa setiap kata untuk memastikan hanya kata benda yang diambil, dengan syarat panjang *lemma* minimal tiga karakter, kemudian dihitung frekuensi kemunculannya. Proses ini berlangsung terus-menerus hingga seluruh kumpulan baris

ulasan selesai dianalisis, lalu daftar kata ini disimpan dan dianggap sebagai kandidat untuk kemudian dipilih sebagai aspek yang relevan.

Kemudian dilakukan kurasi terhadap daftar kandidat aspek dengan intervensi manual, yaitu meninjau daftar kata benda yang paling sering muncul. Setiap istilah dilihat satu per satu untuk dipilih, sedangkan istilah yang tidak berkaitan secara langsung dihilangkan dari daftar. Hasil dari proses ini adalah kamus aspek final yang bersih dan sesuai konteks untuk mempermudah interpretasi di tahap berikutnya.

Dengan kamus aspek yang telah siap, proses dilanjutkan ke tahap implementasi *rule-based* dengan cara pencocokan sederhana ke seluruh *dataset* ulasan. Sistem akan memeriksa apakah *token* tertentu dari ulasan tersebut terdapat dalam kamus aspek. Jika ditemukan kecocokan, *token* tersebut akan dicatat sebagai aspek yang terdeteksi dalam ulasan tersebut.

Tabel 3. Ekstraksi Aspek

cleaned review	detected aspect
pas tangan ukur xl tentu nyaman motor karena jari bisa tap layar hp makasih	pas, ukur
makasih min sudah sampai kirim cepat hari sudah sampai beda pulau dapat harga	kirim, harga, awet
diskon sekitar sepatu bagus nyesal beli moga awet	
kualitas sudah ragu lagi produk original penjual ramah gercep	kualitas, original
simple keren model makasih seller tokped kurir	model, kurir
lingkar kepala cocok nyaman pakai produk kualitas penjual responsif kirim cukup lama	kualitas, kirim

Fondasi dari setiap model *supervised learning* adalah data latih yang berkualitas. Maka, langkah pertama dalam tahap ini adalah melakukan proses anotasi atau pelabelan data secara manual. Potongan data akan diambil dari *dataset* total hasil ekstraksi aspek untuk dijadikan sampel, targetnya kurang lebih 20% dari total *dataset* untuk menghasilkan data yang representatif. Setiap baris data dalam sampel ini akan dibaca teks ulasannya untuk dipahami konteksnya, lalu diberikan label sentimen yang sesuai pada setiap aspek yang terdeteksi. Proses ini akan dijadikan acuan oleh model untuk belajar.

Tabel 4. Anotasi Data

cleaned review	detected aspect	sentiment
pas tangan ukur xl tentu nyaman motor karena jari bisa tap layar hp	pas,ukur	positif
makasih		
makasih min sudah sampai kirim cepat hari sudah sampai beda pulau dapat	kirim,harga,awet	positif
harga diskon sekitar sepatu bagus nyesal beli moga awet		
kualitas sudah ragu lagi produk original penjual ramah gercep	kualitas,original	positif
simple keren model makasih seller tokped kurir	model,kurir	netral
lingkar kepala cocok nyaman pakai produk kualitas penjual responsif kirim cukup lama	kualitas,kirim	Positif

Setelah proses anotasi, *dataset* dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Data latih digunakan agar model dapat mempelajari pola, sementara data uji dipakai untuk mengevaluasi kinerjanya pada data baru yang belum pernah dilihat. Pembagian ini umumnya dilakukan dengan rasio 80% untuk data latih dan 20% untuk data uji.

E. Pelatihan Model

Untuk memungkinkan algoritma *K-Nearest Neighbor (KNN)* memproses data teks, data tersebut harus diubah menjadi format numerik melalui proses vektorisasi. Salah satu metode pembobotan kata yang populer untuk tujuan ini adalah *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini bekerja dengan memberikan bobot pada setiap kata untuk merefleksikan tingkat kepentingannya dalam sebuah dokumen relatif terhadap kumpulan dokumen secara keseluruhan.

Bobot dari kata ini dihitung dari dua komponen utama. Pertama adalah *Term Frequency (TF)*, yang mengukur frekuensi kemunculan sebuah kata di dalam satu dokumen. Kedua adalah *Inverse Document Frequency (IDF)*, yang mengukur keunikan kata tersebut dengan menghitung seberapa jarang ia muncul di seluruh kumpulan dokumen. Dengan demikian, kata yang spesifik dan informatif akan mendapatkan bobot *IDF* yang tinggi, sementara kata umum yang sering muncul di banyak dokumen akan memiliki bobot rendah. Perhitungan *TF-IDF* dapat dilakukan menggunakan persamaan sebagai berikut:

$$tfidf(t, d) = tf(t, d) \times idf(t) \text{ dengan } idf(t) = \log \frac{N}{|\{d' : t \in d'\}|} \quad (1)$$

Keterangan:

- $TF * IDF (d, t)$: Bobot *Term Frequency-Inverse Document Frequency*
- $TF (d, t)$: jumlah munculnya *term t* pada dokumen *d*
- N : total dokumen
- $df(t)$: jumlah dokumen yang terdapat *term t*

Hasil akhir dari proses *TF-IDF* adalah sebuah matriks yang didapat dari perkalian nilai *TF* dan *IDF*. Melalui cara ini, setiap ulasan diubah menjadi representasi vektor numerik yang siap digunakan untuk tahap pelatihan model.

Setelah data ulasan melalui proses vektorisasi, kemudian diterapkan algoritma klasifikasi *K-Nearest Neighbors (KNN)*. *KNN* adalah algoritma *supervised learning* yang tergolong dalam *instance-based learning* dan bekerja berdasarkan prinsip kedekatan jarak. Metode ini menyimpan seluruh data latih untuk dijadikan pembandingan saat ada data baru yang akan diklasifikasikan.

Prinsip kerjanya adalah dengan menghitung jarak dari satu data uji ke seluruh data dalam himpunan data latih. Setelah jarak dihitung, kelas untuk data uji tersebut akan ditentukan berdasarkan kelas mayoritas dari sejumlah tetangga terdekatnya (nilai K). Dengan kata lain, sebuah data cenderung memiliki kelas yang sama dengan data-data lain yang paling dekat dengannya di dalam ruang fitur. Langkah-langkahnya adalah sebagai berikut:

1. Penentuan Nilai K
Menentukan jumlah tetangga terdekat (K) yang akan digunakan sebagai acuan untuk proses *voting*. Nilai K ini biasanya merupakan bilangan ganjil untuk menghindari hasil seri, misalnya $K = 3$. Artinya, akan mencari 3 tetangga terdekat dari data uji.
2. Perhitungan Jarak
Perhitungan ini melihat jarak atau kemiripan antara data baru yang akan diklasifikasikan dengan semua data yang ada di dalam data latih. Untuk data teks yang telah direpresentasikan sebagai vektor *TF-IDF*, metrik yang digunakan adalah *Cosine Similarity*, yang mengukur kosinus sudut antara dua vektor. Semakin kecil sudutnya, semakin tinggi nilai kemiripannya.
Rumus *Cosine Similarity*:
$$\text{sim}(d_i, d_j) = \frac{v_{d_i} \cdot v_{d_j}}{\|v_{d_i}\| \|v_{d_j}\|} \quad (2)$$
di mana V_d adalah vektor fitur dari dokumen d dan simbol \cdot menyatakan perkalian antar vektor.
3. Identifikasi Tetangga Terdekat
Setelah menghitung jarak, kemudian mengurutkan semua data latih berdasarkan jaraknya dari data baru, lalu memilih K data dengan jarak terdekat atau kemiripan tertinggi.
4. Voting
Melakukan *voting* di antara K tetangga terdekat tersebut. Label kelas yang paling banyak muncul di antara K tetangga akan ditetapkan sebagai label prediksi untuk data baru.

F. Evaluasi dan Implementasi Model

Seluruh proses pelatihan model dijalankan dengan menggunakan kode *Python* dengan bantuan *library* pendukung *scikit-learn*. Alur *pipeline* pelatihan hasilnya disimpan menjadi sebuah file dengan format *joblib*.

Tujuan utama dari tahap evaluasi ini adalah untuk menilai kemampuan generalisasi model, yaitu seberapa akurat kemampuannya dalam memberikan prediksi pada data baru yang belum pernah dilihat. Prosesnya dilakukan dengan menggunakan data uji yang telah dipisahkan sebelumnya. *Pipeline* model yang tersimpan dalam file *joblib* digunakan untuk membuat prediksi sentimen pada setiap baris data uji.

Hasil prediksi tersebut kemudian dibandingkan secara sistematis dengan label sentimen manual yang sebenarnya. Untuk memvisualisasikan dan mengukur kinerja ini, digunakan *Confusion Matrix* yang menyajikan rincian prediksi benar dan salah untuk setiap kelas, serta menjadi dasar untuk menghitung metrik evaluasi lainnya. Untuk menghitung akurasi, presisi, dan *recall*, digunakan rumus sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = 2 \times \frac{TP}{TP+FN} \quad (5)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Keterangan:

- TP (*True Positive*) : jumlah data positif yang berhasil diklasifikasikan dengan benar sebagai positif
- TN (*True Negative*) : jumlah data negatif yang diklasifikasikan benar sebagai negatif
- FP (*False Positive*) : jumlah data negatif yang keliru diprediksi sebagai positif
- FN (*False Negative*) : jumlah data positif yang keliru diprediksi sebagai negatif

Setelah model divalidasi dan kinerjanya dinilai memuaskan berdasarkan hasil evaluasi pada data uji, langkah terakhir adalah mengimplementasikan model tersebut pada skala penuh. Pada tahap ini, model yang sama dipakai kembali, namun bukan pada data uji yang kecil. Model ini diterapkan pada keseluruhan *dataset* yang telah melalui tahap ekstraksi aspek untuk memprediksi label sentimen (*Positif*, *Negatif*, atau *Netral*) pada setiap aspek yang terdeteksi di setiap ulasan dalam *dataset*.

Hasil akhir dari tahap ini adalah *dataset final* yang lengkap, berisi teks ulasan, aspek terdeteksi, dan prediksi sentimen untuk setiap aspek. *Dataset* ini akan menjadi fondasi utama untuk analisis, visualisasi, dan penarikan kesimpulan.

Tabel 5. Hasil Prediksi Sentimen Model

cleaned_review	detected_aspect	sentiment	predicted_sentiment
pas tangan ukur xl tentu nyaman motor karena jari bisa tap layar hp makasih	pas,ukur	positif	positif
makasih min sudah sampai kirim cepat hari sudah sampai beda pulau dapat harga	kirim,harga,awet	positif	positif
diskon sekitar sepatu bagus nyesal beli moga awet			
kualitas sudah ragu lagi produk original penjual ramah gercep	kualitas,original	positif	positif
simple keren model makasih seller tokped kurir	model,kurir	netral	netral
lingkar kepala cocok nyaman pakai produk kualitas penjual responsif kirim cukup lama	kualitas,kirim	positif	positif

G. Perancangan Sistem

Untuk menunjukkan kegunaan dari alur kerja analisis sentimen yang telah dilakukan, proses di dalam penelitian ini akan dikembangkan dalam bentuk sebuah sistem sederhana. Tujuan tahap ini adalah untuk membangun sebuah alat yang mampu menjalankan keseluruhan proses *Aspect-Based Sentiment Analysis (ABSA)* secara terintegrasi. Sistem dirancang untuk penggunaan sekali pakai, di mana pengguna mengunggah *dataset* dan sistem langsung memprosesnya untuk menampilkan hasil, tanpa memerlukan *database* untuk penyimpanan data.

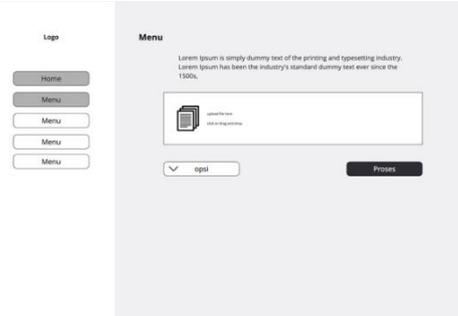
1. Arsitektur Sistem

Arsitektur sistem yang akan dibangun menggunakan pendekatan *client-server*. Pengguna akan berinteraksi melalui antarmuka *web* yang mudah dipahami. Ketika pengguna mengunggah file ulasan dan memulai analisis, permintaan akan dikirim ke *backend* yang berisi seluruh logika pemrosesan seperti yang dilakukan pada proses analisis sentimen sebelumnya. *Backend* kemudian akan menjalankan *pipeline ABSA* mulai dari *pre-processing* hingga klasifikasi sentimen, dan mengirimkan kembali hasil analisis yang sudah terstruktur ke *frontend* untuk divisualisasikan.

2. Desain Antarmuka Pengguna

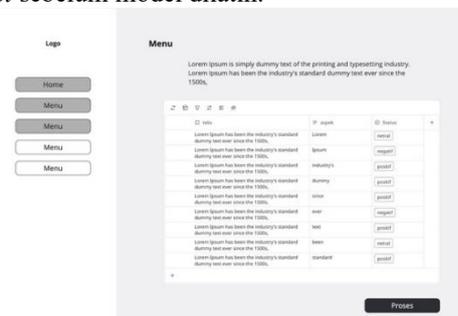
Desain antarmuka pengguna dirancang dengan mempertimbangkan kemudahan penggunaan, dan akan dibangun sebagai *Single Page Application (SPA)* untuk memberikan pengalaman pengguna yang lebih cepat dan efisien. Alur kerja sistem secara visual terbagi menjadi tiga tampilan utama: halaman unggah data, halaman pelabelan anotasi manual, dan halaman *dashboard* hasil analisis.

Pertama, pengguna disambut oleh tampilan awal yang minimalis, seperti terlihat pada Gambar 3. Halaman ini fokus pada satu aksi utama, yaitu menyediakan komponen untuk mengunggah file ulasan, memilih kolom data yang relevan, dan menekan tombol “Mulai Proses” untuk memulai analisis.



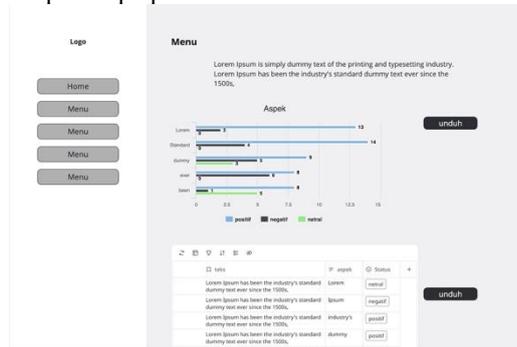
Gambar 3. Wireframe Halaman Unggah Data

Selanjutnya, untuk kebutuhan penyiapan data latih yang berkualitas, sistem menyediakan antarmuka pelabelan manual yang krusial. Sebagaimana diilustrasikan pada Gambar 4, antarmuka ini menampilkan sampel ulasan beserta aspeknya, di mana peneliti dapat menetapkan label sentimen (*positif*, *negatif*, atau *netral*) untuk memastikan kualitas *dataset* sebelum model dilatih.



Gambar 4. Wireframe Halaman Pelabelan Manual

Terakhir, setelah model diterapkan, seluruh wawasan disajikan dalam *dashboard* hasil analisis yang ringkas, sesuai rancangan pada Gambar 5. *Dashboard* ini memvisualisasikan temuan utama seperti diagram distribusi sentimen per aspek serta ringkasan kekuatan dan kelemahan produk, sehingga pengguna dapat memahami hasil analisis secara cepat tanpa perlu memeriksa data mentah.



Gambar 5. Wireframe Halaman Dashboard Hasil

H. Pengujian Sistem

Pengujian sistem dilakukan untuk memverifikasi bahwa setiap komponen fungsional berjalan sesuai spesifikasi, memastikan sistem bebas dari *error*, dan mampu menghasilkan *output* yang akurat untuk berbagai skenario *input*.

Untuk mencapai tujuan ini, metode pengujian yang digunakan adalah *White Box Testing*. Pendekatan ini dipilih karena memungkinkan pengujian dilakukan dengan melihat langsung struktur kode internal dan alur logika di dalam sistem. Fokus utamanya adalah memvalidasi proses algoritma dengan merancang serangkaian kasus uji yang mencakup penggunaan normal, kasus-kasus khusus, hingga potensi kegagalan. Dengan memeriksa jalur-jalur logika secara internal, pengujian ini memberikan keyakinan yang lebih mendalam terhadap keandalan dan kebenaran alur kerja sistem secara keseluruhan.

3. RESULTS AND DISCUSSION

A. Hasil Ekstraksi Aspek

Tahap ekstraksi aspek berhasil mengidentifikasi berbagai kategori fitur produk dan layanan yang dibicarakan oleh pelanggan seperti yang disajikan pada Tabel 6.

Tabel 6. Distribusi Aspek Terdeteksi

Kategori Aspek	Jumlah Kemunculan
Pengiriman	829
Model	765
Kualitas	747
Kualitas;Model	685
Ukuran	539
Kualitas;Ukuran	435
Model;Ukuran	431
Kualitas;Model;Ukuran	423
Kualitas;Pengiriman	405
Pengiriman;Model	363

Data ulasan pelanggan menyoroti tiga kategori dominan: Pengiriman, Model, dan Kualitas. Temuan signifikan lainnya adalah adanya ulasan gabungan (misalnya, "*Kualitas;Model*"), yang mengindikasikan bahwa konsumen cenderung melakukan evaluasi multidimensional. Hal ini menegaskan bahwa bagi pelanggan Eiger, atribut fungsional dan performa produk merupakan faktor pertimbangan utama, selaras dengan citra merek yang mengedepankan durabilitas.

B. Hasil Analisis Sentimen per Aspek

Setelah aspek-aspek utama teridentifikasi, analisis dilanjutkan dengan klasifikasi sentimen untuk memahami persepsi pelanggan. Sebuah model digunakan untuk menentukan apakah opini pada setiap aspek bersifat *positif* atau *negatif*. Hasil yang disajikan dalam Tabel 7 menunjukkan secara jelas apa yang menjadi keunggulan kompetitif dan area perbaikan bagi produk Eiger.

Tabel 7. Distribusi Sentimen Tiap Aspek

Kategori Aspek	Positif	Negatif	Netral
Pengiriman	152	56	621
Model	258	14	493
Kualitas	326	14	407
Kualitas;Model	393	13	279
Ukuran	102	40	397
Kualitas;Ukuran	174	29	232
Model;Ukuran	204	4	223

Hasil pemetaan sentimen menunjukkan bahwa kekuatan utama Eiger di mata pelanggan adalah Kualitas dan Model produk. Kedua aspek ini, terutama saat digabungkan, secara konsisten menerima sentimen positif yang sangat tinggi, mengonfirmasi citra mereknya yang andal dan modern. Adapun tantangan terbesar Eiger adalah pada aspek Pengiriman, yang menjadi sumber keluhan negatif paling banyak. Selain itu, masalah Ukuran produk juga menjadi sumber sentimen negatif yang signifikan dan memerlukan perhatian.

Temuan menarik adalah banyaknya ulasan netral, terutama pada aspek *Pengiriman*. Hal ini menunjukkan bahwa sebagian besar pelanggan hanya memberikan ulasan informatif (misalnya “*paket sudah sampai*”), yang bukan merupakan pujian ataupun keluhan, namun tetap bisa menjadi data berharga, misalnya untuk mengukur keberhasilan transaksi.

C. Evaluasi Kinerja Klasifikasi Sentimen

Tahap selanjutnya adalah mengevaluasi kinerja algoritma *K-Nearest Neighbor (KNN)* dalam mengklasifikasikan sentimen. Evaluasi ini menggunakan 268 data uji yang telah dilabeli secara manual untuk membuktikan bahwa sistem yang dibangun akurat. Kinerja model diukur dan ditampilkan melalui *Confusion Matrix* beserta metrik performa yang dihitung darinya.

Tabel 8. Confusion Matrix

	Prediksi: Negatif	Prediksi: Netral	Prediksi: Positif
Aktual: Negatif	13	10	11
Aktual: Netral	4	60	30
Aktual: Positif	2	46	92

Tabel 9. Performa Model

	Precision	Recall	F1-score	Support
negatif	0.684	0.382	0.491	34
netral	0.517	0.638	0.571	94
positif	0.692	0.657	0.674	140
accuracy			0.616	268
macro avg	0.631	0.559	0.579	268
weighted avg	0.63	0.616	0.615	268

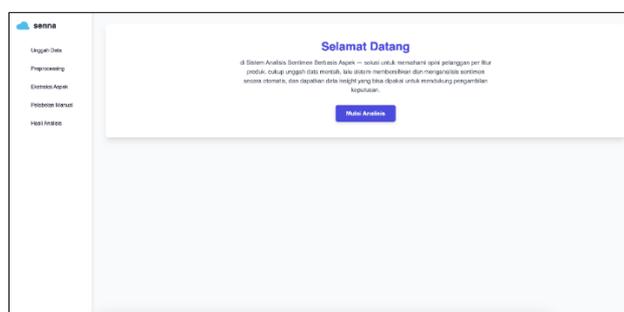
Model klasifikasi sentimen ini mencapai akurasi keseluruhan 61,6%, namun kinerjanya tidak merata. Model sangat baik dalam mengenali sentimen *positif* (*F1-score* 0.674), tetapi sangat lemah dalam mendeteksi sentimen *negatif* (gagal mengidentifikasi sekitar 62% kasus) dan sering salah mengklasifikasikannya sebagai *positif* atau *netral*. Meskipun kinerjanya (*F1-score* rata-rata 0.579) cukup sebagai bukti konsep awal (*proof-of-concept*), model ini memiliki bias yang jelas terhadap kelas *positif* dan kesulitan membedakan sentimen lainnya. Oleh karena itu, ada ruang yang signifikan untuk perbaikan, terutama dalam meningkatkan deteksi sentimen *negatif*, misalnya dengan menangani data yang tidak seimbang atau menggunakan algoritma yang lebih kompleks.

D. Tampilan Antarmuka Sistem

Sebagai hasil akhir penelitian, sebuah sistem dengan antarmuka pengguna sederhana telah dibangun. Bagian ini akan mendemonstrasikan alur kerja sistem tersebut melalui serangkaian tangkapan layar, untuk menunjukkan bagaimana hasil analisis disajikan secara visual dan interaktif.

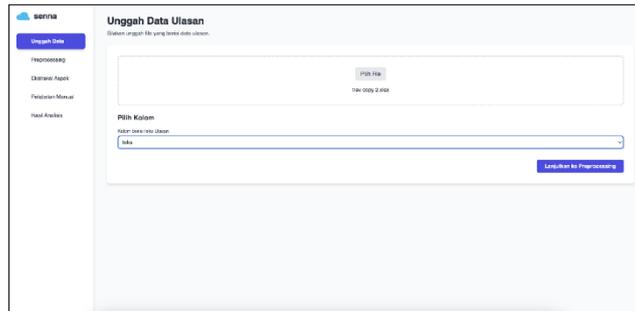
1. Halaman Utama dan Proses Unggah Data

Proses analisis diawali pada halaman utama sistem, seperti yang terlihat pada Gambar 6. Halaman ini berfungsi sebagai *landing page* yang menyambut pengguna dan memberikan deskripsi singkat mengenai fungsi utama dari prototipe.



Gambar 6. Landing Page

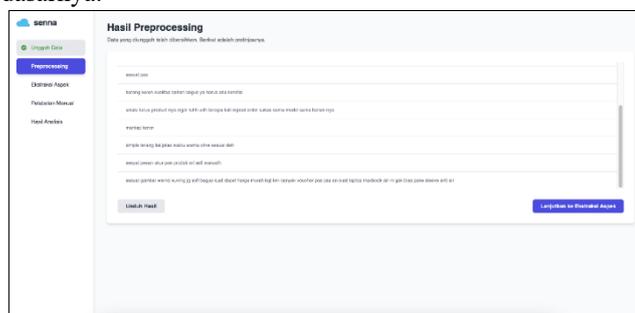
Untuk memulai analisis, pengguna perlu mengunggah *dataset* ulasan. Seperti yang ditunjukkan pada Gambar 7, sistem menyediakan tombol “*Upload File*”. Setelah diklik, pengguna dapat memilih file ulasan dari perangkatnya.



Gambar 7. Halaman Unggah Data Ulasan

2. Tampilan Hasil Pra-pemrosesan Teks

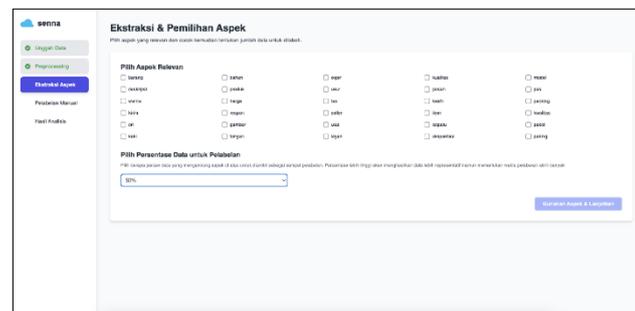
Setelah data diunggah, sistem secara otomatis melakukan *pre-processing*. Sistem kemudian menampilkan sampel data “*sebelum dan sesudah*” dibersihkan seperti pada Gambar 8, agar pengguna dapat memverifikasi bahwa elemen *noise* (tanda baca, kapitalisasi, *stopwords*) telah berhasil dihilangkan dan teks sudah diubah ke bentuk dasarnya.



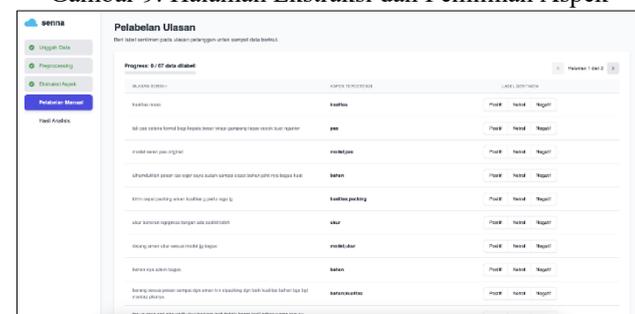
Gambar 8. Halaman Hasil Preprocessing

3. Tampilan Proses Ekstraksi dan Pelabelan Aspek

Pada tahap selanjutnya, sistem secara otomatis mengekstraksi kandidat aspek dari ulasan, yang kemudian dapat ditinjau oleh pengguna melalui sebuah antarmuka seperti pada Gambar 9. Sistem juga memfasilitasi proses kunci dalam *supervised learning*: pengguna dapat secara manual memberi label sentimen (*Positif, Negatif, atau Netral*) pada setiap kalimat untuk menyiapkan data latih dan uji, seperti ditunjukkan pada Gambar 10.



Gambar 9. Halaman Ekstraksi dan Pemilihan Aspek

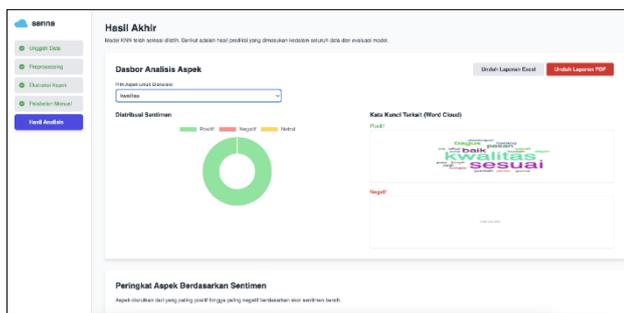


Gambar 10. Halaman Pelabelan Ulasan

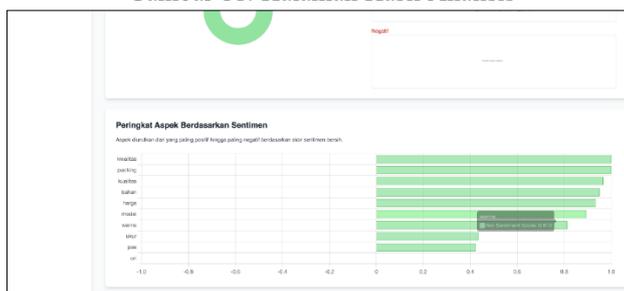
4. Halaman Hasil Akhir dan Visualisasi

Setelah analisis selesai, sistem akan menampilkan *dashboard* hasil akhir seperti terlihat pada Gambar 11 dan 12. *Dashboard* ini merangkum temuan utama secara visual dengan menyoroti dua hal: daftar aspek

yang paling dominan dibicarakan dan grafik batang yang membandingkan distribusi sentimen untuk setiap aspek kunci tersebut.



Gambar 11. Halaman Hasil Analisis



Gambar 12. Halaman Hasil Analisis

E. Pengujian Sistem

Setelah demonstrasi antarmuka, tahap selanjutnya adalah pengujian sistem. Serangkaian kasus uji mandiri dirancang untuk mencakup skenario penggunaan normal maupun pengecualian (*exception*). Tujuannya adalah untuk memverifikasi alur kerja internal kode dan memastikan setiap fitur berfungsi sesuai spesifikasi rancangan, dengan hasil yang disajikan pada Tabel 10.

Tabel 10. Hasil Whitebox Testing

No.	Kasus Uji	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian
1	Membersihkan Teks	"Kualitas BAGUS!!, harganya juga oke."	"kualitas bagus harga oke"	"kualitas bagus harga oke"
2	Menghapus Stopword	"ukurannya pas tapi pengirimannya lama"	"ukuran pas pengiriman lama"	"ukuran pas pengiriman lama"
3	Menangani Kata Slang	"Kualitasnya bgs bgt."	Kata slang tetap ada	"kualitas bgs bgt"
4	Mengekstrak Kata Benda	"desain tasnya modern dan bahannya kuat"	['desain', 'tas', 'bahan']	['desain', 'tas', 'bahan']
5	Gagal Deteksi Aspek Implisit	"baru dipakai sekali sudah sobek"	[]	[]
6	Menjalankan Pipeline ABSA	DataFrame valid dengan teks ulasan dan aspek	Pipeline berjalan tanpa error	Pipeline berjalan tanpa error

Berdasarkan pengujian tersebut, terlihat bahwa sistem telah mampu menangani skenario utama dengan baik. Fungsi pembersihan teks pada Skenario 1 dan 2 berhasil menghapus karakter tidak penting serta *stopword* dari kalimat, menghasilkan *output* yang sesuai dengan ekspektasi. Ekstraksi kata benda pada Skenario 4 juga berhasil mengidentifikasi semua aspek eksplisit yang terkandung dalam kalimat.

Sementara itu, fungsi menangani *slang* (Skenario 3) dan deteksi aspek implisit (Skenario 5) menunjukkan keterbatasan. Sistem belum mampu menormalkan kata *slang* seperti "bgs" atau "bgt" karena tidak dilengkapi dengan kamus *slang* Bahasa Indonesia. Selain itu, sistem gagal mengidentifikasi aspek implisit seperti "kualitas" dalam kalimat keluhan "sudah sobek", karena metode yang digunakan berbasis *POS Tagging* untuk kata benda eksplisit.

Pipeline keseluruhan yang diuji dalam Skenario 6 mampu dijalankan tanpa adanya *error*, membuktikan bahwa integrasi antar komponen sistem sudah berjalan secara konsisten dan stabil.

F. Pembahasan

Setelah menyajikan data kuantitatif, sub-bab ini akan membahas makna yang lebih dalam dari temuan tersebut. Pembahasan akan berfokus pada analisis kekuatan dan kelemahan produk Eiger, memberikan saran praktis bagi perusahaan, serta menguraikan keterbatasan penelitian dan arah pengembangan selanjutnya. Intinya, bagian ini bertujuan untuk menerjemahkan data menjadi wawasan yang dapat ditindaklanjuti.

1. Temuan Utama

Analisis sentimen berbasis aspek ini memberikan wawasan yang jauh lebih mendalam daripada sekadar peringkat bintang, dengan secara spesifik memetakan nilai jual utama dan kelemahan dalam pengalaman pelanggan Eiger. Kekuatan utama Eiger terletak pada *Kualitas* dan *Model* produknya. Sentimen positif yang sangat dominan pada kedua aspek ini membuktikan bahwa Eiger telah berhasil memenuhi ekspektasi pelanggan akan durabilitas dan desain, serta citra mereknya selaras dengan pengalaman nyata pengguna.

Sedangkan kelemahan terbesar justru datang dari aspek non-produk, yaitu *Pengiriman* dan *Ukuran*. Masalah pada logistik dan ketidakjelasan informasi *size chart* terbukti dapat merusak pengalaman belanja *e-commerce* secara keseluruhan, meskipun kualitas produknya sendiri sangat baik. Terkait *Harga*, meskipun ada sentimen negatif, hal ini diimbangi oleh sentimen yang sangat positif pada *Kualitas*. Ini mengindikasikan bahwa sebagian besar pelanggan merasa harga premium Eiger sepadan dengan kualitas yang didapatkan, namun tetap ada segmen pasar yang sensitif terhadap harga.

2. Implikasi Praktis

Temuan dari analisis ini dapat diterjemahkan menjadi beberapa saran strategis yang bisa ditindaklanjuti oleh Eiger:

- a) Untuk Pemasaran & Komunikasi: Tim pemasaran disarankan untuk memanfaatkan testimoni pelanggan asli yang secara spesifik memuji kualitas dan desain produk. Menggunakan *bukti sosial* ini akan membuat materi promosi terasa lebih otentik dan meyakinkan bagi calon pembeli.
- b) Untuk Operasional & Layanan Pelanggan: Tingginya keluhan pada aspek *Pengiriman* adalah sinyal penting. Perusahaan perlu melakukan audit terhadap mitra logistik mereka, mengevaluasi proses pemenuhan pesanan, dan meningkatkan transparansi pelacakan untuk mengelola ekspektasi pelanggan dengan lebih baik.
- c) Untuk Pengembangan Produk & UX: Masalah pada aspek *Ukuran* memberikan masukan langsung untuk perbaikan. Solusinya adalah dengan menstandarisasi panduan ukuran di semua platform *e-commerce* dan menambahkannya dengan referensi yang lebih interaktif atau visual (seperti foto *size chart* di halaman produk) untuk mengurangi kebingungan dan tingkat pengembalian barang.

4. CONCLUSION

Berdasarkan analisis 3.670 ulasan unik, kesimpulan utama menunjukkan bahwa pelanggan sangat puas dengan kualitas produk Eiger namun sering kecewakan oleh pengalaman transaksionalnya. Fokus utama pelanggan adalah pada *kualitas*, *harga*, dan *bahan*, di mana kekuatan Eiger terbukti terletak pada atribut produknya sendiri (seperti desain, kualitas, dan kenyamanan yang mendapat sentimen positif sangat tinggi). Sebaliknya, kelemahan terbesar ada pada aspek layanan seperti *pengiriman* yang menjadi sumber keluhan utama, diikuti oleh isu *harga* dan *ukuran*. Walaupun analisis ini dilakukan menggunakan sistem bukti konsep dengan akurasi 61,6% yang memiliki keterbatasan, temuannya secara jelas mengarahkan pada kesimpulan strategis: untuk meningkatkan kepuasan secara menyeluruh, fokus perbaikan Eiger harus diperluas dari sekadar produk ke ekosistem layanannya.

Penelitian ini memiliki beberapa keterbatasan utama yang perlu diakui. Secara metodologis, metode ekstraksi aspek hanya mampu menangkap topik yang disebutkan secara eksplisit (bukan implisit), dan model analisisnya tidak dirancang untuk dapat memahami bahasa kompleks seperti sarkasme atau ironi. Selain itu, keterbatasan juga datang dari sumber data yang terbatas hanya pada satu platform *e-commerce* (Tokopedia), sehingga hasilnya mungkin tidak sepenuhnya mewakili seluruh basis pelanggan. Terakhir, evaluasi pada proses ekstraksi aspek itu sendiri dilakukan secara kualitatif, bukan melalui pengukuran kuantitatif formal.

Berdasarkan temuan dan keterbatasan yang ada, penelitian ini menghasilkan beberapa saran yang dapat ditindaklanjuti. Secara praktis, Eiger disarankan untuk memanfaatkan ulasan positif tentang kualitas dan desain sebagai testimoni pemasaran, melakukan audit logistik untuk mengatasi masalah pengiriman, dan menstandarisasi panduan ukuran untuk mengurangi kebingungan pelanggan. Sementara itu, untuk penelitian selanjutnya, disarankan untuk mengembangkan metode yang lebih canggih (seperti *deep learning* dan model *transformer*) untuk meningkatkan akurasi analisis, serta memperluas cakupan riset dengan menyertakan data kompetitor dari berbagai platform untuk mendapatkan pemahaman pasar yang lebih komprehensif.

REFERENCES

- Afonso, A. P., Carneiro, J., & Azevedo, A. I. (2023). The impact of COVID-19 on e-commerce: A systematic review of the literature on the purchasing behavior of online retail consumers. *IBIMA Business Review*, 2024, 1–9. <https://doi.org/10.5171/2024.403212>
- Amalia, P. R., & Winarko, E. (2021). Aspect-based sentiment analysis on Indonesian restaurant review using a combination of convolutional neural network and contextualized word embedding. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 285–295. <https://doi.org/10.22146/ijccs.67306>

- Camilleri, M., & Filieri, R. (2023). Customer satisfaction and loyalty with online consumer reviews: Factors affecting revisit intentions. *International Journal of Hospitality Management*. Advance online publication. <https://doi.org/10.1016/j.ijhm.2023.103575>
- Changchit, C., & Klaus, T. (2020). Determinants and impact of online reviews on product satisfaction. *Journal of Internet Commerce*, 19(2), 102–182. <https://doi.org/10.1080/15332861.2019.1672135>
- Chris, E., Olusegun, J., Favor, B., & Ramon, T. (2024). Aspect-based sentiment analysis of customer reviews in e-commerce. [Conference paper, December].
- Dewi, G. D. P., & Lusikooy, A. E. (2024). E-commerce transformation in Indonesia. *Nation State: Journal of International Studies*, 6(2), 117–138. <https://doi.org/10.24076/nsjis.v6i2.1304>
- Imron, S., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2023). Aspect-based sentiment analysis marketplace product reviews using BERT, LSTM, and CNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3), 586–591. <https://doi.org/10.29207/resti.v7i3.4751>
- Jazuli, A., Widowati, & Kusumaningrum, R. (2023). Aspect-based sentiment analysis on student reviews using the IndoBERT base model. *E3S Web of Conferences*, 448, 02004. <https://doi.org/10.1051/e3sconf/202344802004>
- Kemp, S. A. S. (2025). *Digital 2025: Indonesia*. DataReportal. <https://datareportal.com/reports/digital-2025-indonesia>
- Natasya, W. D. P., Musa, M. I., Haeruddin, M. I. W., Hasbiah, S., & Aslam, N. F. (2024). Influence of online customer review and online customer rating on purchasing decision on Shopee e-commerce. *Journal of Business Management and Economic Development*, 2(2), 740–750. <https://doi.org/10.59653/jbmed.v2i02.720>
- Rachmiani, R., Oktadinda, N. K., & Fauzan, T. R. (2023). The impact of online reviews on consumer purchasing decisions. *International Journal of Early Childhood Special Education*, 14(5), 504–515. <https://doi.org/10.48047/intjecse/v14i5.1135>
- Shankar, V., Kalyanam, K., Setia, P., Golmohammadi, A., Tirunillai, S., Douglass, T., Hennessey, J., Bull, J. S., & Waddoups, R. (2021). How technology is changing retail. *Journal of Retailing*, 97(1), 13–27. <https://doi.org/10.1016/j.jretai.2020.10.006>
- Shukla, D., & Dwivedi, S. K. (2025). Sentiment analysis versus aspect-based sentiment analysis versus emotion analysis from text: A comparative study. *International Journal of System Assurance Engineering and Management*, 16(2), 512–531. <https://doi.org/10.1007/s13198-024-02666-4>
- Xu, X., Wang, Y., Zhu, Q., & Zhuang, Y. (2024). Time matters: Investigating the asymmetric reflection of online reviews on customer satisfaction and recommendation across temporal lenses. *International Journal of Information Management*, 75, 102733. <https://doi.org/10.1016/j.ijinfomgt.2023.102733>
- Ziegele, M., & Weber, M. (2015). Example, please! Comparing the effects of single customer reviews and aggregate review scores on online shoppers' product evaluations. *Journal of Consumer Behaviour*, 14(2), 103–114. <https://doi.org/10.1002/cb.1503>