

Analisis Sentimen Ulasan Pelanggan pada Platform E-Commerce Menggunakan IndoBERT untuk Mendukung Evaluasi Kualitas Produk

M. Rifai Susanto^{*1}, Angelina Hadriani², Maman Nur Saman³

^{*1,3}Bisnis Digital, Politeknik Bisnis Digital Indonesia, Bogor

²Sistem Komputer, Universitas Pamulang, Banten

e-mail: ^{*1}rifai.temprina@gmail.com, ²angelinahadriani03234@unpam.ac.id, ³mamannursaman98@gmail.com

Abstrak

Ulasan pelanggan pada platform e-commerce dapat dimanfaatkan untuk memahami persepsi konsumen terhadap kualitas produk. Penelitian ini bertujuan menganalisis sentimen ulasan pelanggan Tokopedia menggunakan IndoBERT serta memanfaatkan hasil klasifikasi untuk mendukung evaluasi kualitas produk. Dataset setelah proses pembersihan terdiri atas 37.180 ulasan dan diberi label berdasarkan rating menjadi sentimen negatif, netral, dan positif. Data dibagi menggunakan stratified sampling dengan proporsi 80% data pelatihan, 10% validasi, dan 10% pengujian. Model IndoBERT dievaluasi menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Hasil pengujian menunjukkan accuracy sebesar 92,44%, weighted F1-score 92,67%, dan macro F1-score 63,36%. Dari 3.718 data pengujian, 92,33% diklasifikasikan positif, 5,70% netral, dan 1,96% negatif. Sentiment Quality Score (SQS) sebesar 0,9037 menunjukkan persepsi pelanggan yang sangat positif. Kategori olahraga memperoleh SQS tertinggi sebesar 0,9424, sedangkan kategori handphone terendah sebesar 0,7634. Hasil ini menunjukkan bahwa IndoBERT efektif untuk klasifikasi sentimen dan dapat mendukung evaluasi kualitas produk berdasarkan persepsi pelanggan.

Kata Kunci: analisis sentimen; IndoBERT; ulasan produk; e-commerce

Abstract

Customer reviews on e-commerce platforms can be utilized to understand consumer perceptions of product quality. This study aims to analyze Tokopedia customer review sentiment using IndoBERT and employ the classification results to support product quality evaluation. After data cleaning, the dataset consisted of 37,180 reviews labeled based on ratings into negative, neutral, and positive sentiments. The data were divided using stratified sampling into 80% training, 10% validation, and 10% testing sets. IndoBERT was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The results show an accuracy of 92.44%, a weighted F1-score of 92.67%, and a macro F1-score of 63.36%. Of 3,718 test reviews, 92.33% were classified as positive, 5.70% as neutral, and 1.96% as negative. The Sentiment Quality Score (SQS) reached 0.9037, indicating highly positive customer perceptions. The sports category achieved the highest SQS of 0.9424, while the mobile phone category obtained the lowest score of 0.7634. These results demonstrate that IndoBERT is effective for sentiment classification and can support product quality evaluation based on customer perceptions.

Keywords: sentiment analysis; IndoBERT; product reviews; e-commerce

I. PENDAHULUAN

Perkembangan e-commerce telah meningkatkan jumlah ulasan pelanggan yang tersedia secara daring. Ulasan tersebut tidak hanya membantu calon konsumen dalam mempertimbangkan keputusan pembelian, tetapi juga dapat dimanfaatkan oleh penjual dan produsen untuk memahami persepsi pelanggan terhadap produk. Informasi yang terkandung dalam ulasan dapat mencerminkan kepuasan, keluhan, kesesuaian produk dengan

deskripsi, serta persepsi terhadap kualitas produk. Oleh karena itu, analisis terhadap ulasan pelanggan dapat memberikan informasi yang berguna untuk mendukung evaluasi dan peningkatan kualitas produk (Daza et al., 2024).

Tokopedia sebagai salah satu platform e-commerce di Indonesia menyediakan fasilitas ulasan bagi pelanggan setelah melakukan pembelian. Besarnya jumlah ulasan menyebabkan proses evaluasi secara manual menjadi kurang efisien.

Selain itu, ulasan berbahasa Indonesia sering mengandung bahasa tidak formal, singkatan, variasi penulisan, dan konteks tertentu yang dapat menyulitkan proses identifikasi sentimen. Analisis sentimen menjadi salah satu pendekatan dalam Natural Language Processing (NLP) yang dapat digunakan untuk mengidentifikasi kecenderungan opini pelanggan secara otomatis ke dalam kategori sentimen tertentu. Singh et al. (2022) menunjukkan bahwa analisis sentimen terhadap ulasan pelanggan e-commerce dapat membantu mengidentifikasi opini konsumen secara lebih sistematis.

Sejumlah penelitian telah menerapkan metode machine learning untuk analisis sentimen ulasan produk Tokopedia. Aulia dan Hermawan (2023) membandingkan Support Vector Machine (SVM), Naïve Bayes, dan Perceptron dalam mengklasifikasikan sentimen ulasan produk Tokopedia. Sementara itu, Alaiya et al. (2025) menerapkan SVM dengan pembobotan TF-IDF dan memperoleh akurasi yang tinggi. Namun, pendekatan berbasis TF-IDF memiliki keterbatasan karena lebih menekankan bobot dan kemunculan kata sehingga belum sepenuhnya menangkap hubungan semantik dan konteks antarkata dalam suatu ulasan.

Perkembangan arsitektur Transformer memberikan alternatif yang lebih baik dalam memahami konteks teks. Untuk bahasa Indonesia, IndoBERT menjadi salah satu model yang relevan karena dikembangkan menggunakan korpus berbahasa Indonesia. Aji dan Aditya (2025) menerapkan pre-trained IndoBERT pada ulasan produk Tokopedia dan memperoleh akurasi serta F1-score sebesar 97%, lebih tinggi dibandingkan model LSTM yang menggunakan FastText, GloVe, dan Word2Vec. Hasil tersebut menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks semantik pada ulasan produk berbahasa Indonesia. Penelitian internasional juga menunjukkan bahwa model berbasis Transformer efektif digunakan untuk analisis sentimen pada ulasan e-commerce, termasuk pada kondisi data yang tidak seimbang dan memiliki struktur teks yang kompleks (Rasappan et al., 2024; Taneja et al., 2024).

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada performa klasifikasi, seperti akurasi, presisi, recall, dan F1-score. Penelitian Aulia dan Hermawan (2023)

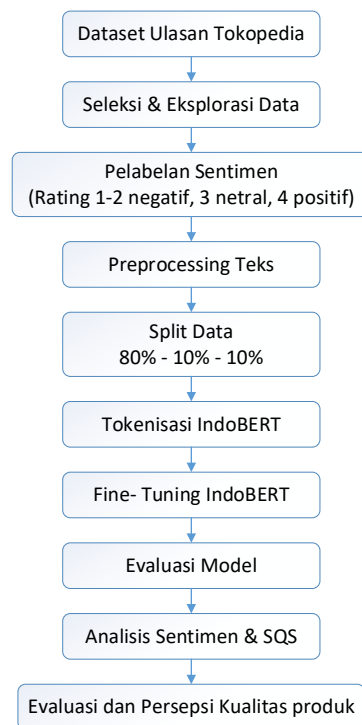
menitikberatkan pada perbandingan algoritma, sedangkan Aji dan Aditya (2025) lebih berfokus pada efektivitas IndoBERT dibandingkan pendekatan LSTM. Dengan demikian, hasil klasifikasi sentimen masih cenderung diposisikan sebagai keluaran akhir dan belum banyak dimanfaatkan secara eksplisit sebagai informasi pendukung evaluasi kualitas produk. Padahal, Daza et al. (2024) menunjukkan bahwa analisis sentimen ulasan e-commerce dapat dimanfaatkan untuk memahami persepsi pelanggan terhadap produk, sedangkan Ray dan Singh (2025) menunjukkan bahwa informasi sentimen dapat dikembangkan lebih lanjut untuk mendukung evaluasi dan rekomendasi produk.

Berdasarkan penelitian terdahulu tersebut, terdapat research gap pada pemanfaatan IndoBERT untuk analisis ulasan produk Tokopedia. Pertama, sejumlah penelitian Tokopedia masih menggunakan algoritma machine learning konvensional dan representasi fitur seperti TF-IDF. Kedua, penelitian yang telah menerapkan IndoBERT lebih banyak berorientasi pada peningkatan performa klasifikasi, sedangkan pemanfaatan hasil sentimen sebagai informasi untuk mendukung evaluasi kualitas produk masih terbatas. Ketiga, studi internasional mulai mengarahkan analisis sentimen pada evaluasi produk, tetapi penerapannya pada ulasan produk berbahasa Indonesia dengan model bahasa khusus Indonesia masih relatif terbatas.

Oleh karena itu, penelitian ini menerapkan IndoBERT untuk mengklasifikasikan sentimen ulasan pelanggan pada platform Tokopedia dan menganalisis hasil klasifikasi sebagai informasi pendukung evaluasi kualitas produk. Kontribusi penelitian tidak hanya terletak pada penerapan model Transformer untuk klasifikasi sentimen berbahasa Indonesia, tetapi juga pada pemanfaatan distribusi dan kecenderungan sentimen pelanggan untuk memberikan gambaran mengenai persepsi terhadap produk. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan menggunakan IndoBERT, mengevaluasi kinerja model berdasarkan metrik klasifikasi, serta menginterpretasikan hasil sentimen sebagai informasi yang dapat mendukung evaluasi kualitas produk pada platform e-commerce.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan memanfaatkan data ulasan pelanggan pada platform Tokopedia. Tahapan penelitian meliputi pengumpulan dan pemahaman data, seleksi atribut, pelabelan sentimen, praproses teks, pembagian dataset, tokenisasi, fine-tuning IndoBERT, evaluasi performa model, serta analisis hasil klasifikasi untuk mendukung evaluasi kualitas produk. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian analisis sentimen menggunakan IndoBERT

Dataset Penelitian

Dataset yang digunakan adalah Tokopedia Product Reviews Dataset, yaitu kumpulan ulasan pelanggan terhadap produk yang dipasarkan pada platform Tokopedia. Atribut utama yang digunakan adalah teks ulasan dan rating, sedangkan atribut nama produk dan kategori dimanfaatkan untuk analisis pada tingkat produk atau kategori apabila tersedia. Sebelum pemodelan dilakukan eksplorasi untuk mengetahui jumlah data, tipe atribut, distribusi rating, nilai kosong, data duplikat, dan panjang teks.

Tabel 1. Atribut dataset yang digunakan

No.	Atribut	Tipe Data	Fungsi
1	Review	Teks	Input model berupa ulasan pelanggan
2	Rating	Numerik	Dasar pembentukan label sentimen
3	Product/Category	Kategorikal	Analisis per produk/kategori
4	Sentiment	Kategorikal	Target klasifikasi

Pelabelan Sentimen

Apabila dataset belum menyediakan label sentimen secara eksplisit, label dibentuk berdasarkan rating pelanggan menggunakan pendekatan weak supervision. Rating 1–2 diberi label negatif, rating 3 diberi label netral, dan rating 4–5 diberi label positif.

Tabel 2. Atribut dataset yang digunakan

Rating	Label Sentimen	Kode
1–2	Negatif	0
3	Netral	1
4–5	Positif	2

$$S(r) = 0 \text{ jika } r \leq 2;$$

$$S(r) = 1 \text{ jika } r = 3;$$

$$S(r) = 2 \text{ jika } r \geq 4$$

Dengan $S(r)$ merupakan label sentimen dan r merupakan rating pelanggan. Pelabelan berbasis rating digunakan sebagai weak supervision sehingga analisis selanjutnya tetap mempertimbangkan kemungkinan ketidaksesuaian antara rating dan isi teks ulasan.

Praproses Data Teks

Praproses dilakukan untuk mengurangi noise tanpa menghilangkan informasi penting yang diperlukan IndoBERT untuk memahami konteks. Tahapan meliputi penghapusan data kosong dan duplikat, URL, tag HTML, karakter tidak relevan, spasi berlebih, serta normalisasi kata tidak baku. Stemming dan penghapusan stopword secara agresif tidak dilakukan karena struktur kalimat dan kata tertentu masih memiliki informasi kontekstual yang penting bagi model Transformer.

Tabel 3. Contoh praproses teks

Tahap	Contoh
Teks awal	Barangnya bagusss bgt!!! Sesuai foto
Cleaning	Barangnya bagusss bgt Sesuai foto
Normalisasi	barangnya bagus banget sesuai foto
Label	Positif

Pembagian Dataset

Dataset yang telah dibersihkan dibagi menjadi data latih, validasi, dan uji menggunakan stratified sampling untuk mempertahankan proporsi masing-masing kelas

Tabel 4. Proporsi pembagian dataset

Dataset	Proporsi	Fungsi
Training	80%	Melatih parameter model
Validation	10%	Memantau pelatihan dan memilih model terbaik
Testing	10%	Menguji kemampuan generalisasi model

$$D = D_{train} \cup D_{val} \cup D_{test} \quad (1)$$

Tokenisasi IndoBERT

Teks ulasan diproses menggunakan tokenizer IndoBERT. Tokenisasi mengubah teks menjadi token yang kemudian dikonversi menjadi input ID serta attention mask. Padding digunakan untuk teks yang lebih pendek daripada panjang maksimum, sedangkan truncation digunakan apabila jumlah token melebihi batas yang ditentukan.

$$x_i = [t_1, t_2, \dots, t_n] \quad (2)$$

$$A_i = [a_1, a_2, \dots, a_n] \quad (3)$$

dengan $a_i = 1$ untuk token aktual dan 0 untuk padding.

Fine-Tuning IndoBERT

Model utama yang digunakan adalah IndoBERT berbasis arsitektur Bidirectional Encoder Representations from Transformers. Representasi masukan diperoleh dari penjumlahan token embedding, segment embedding, dan position embedding. Mekanisme self-attention menghitung hubungan kontekstual antartoken menggunakan Query (Q), Key (K), dan Value (V). Representasi token [CLS] dari lapisan terakhir diteruskan ke lapisan klasifikasi tiga kelas.

$$E = E_{token} + E_{segment} + E_{position} \quad (4)$$

$$Attention(Q, K, V) = softmax((QK^T)/\sqrt{d_k})V \quad (5)$$

$$P(y = c|x) = \frac{e^{z_{c,x}}}{\sum_j e^{z_{j,x}}} \quad (6)$$

$$\hat{y} = \underset{c}{\operatorname{argmax}} P(y = c|x) \quad (7)$$

Fungsi Loss

Pelatihan menggunakan cross-entropy loss. Karena distribusi kelas tidak seimbang, class weighting diterapkan agar kesalahan pada kelas minoritas memperoleh penalti yang lebih besar.

$$L = -\left(\frac{1}{N}\right) \sum_i \sum_c y_i c \log(\hat{y}_i c) \quad (8)$$

Parameter Pelatihan

Tabel 5. Parameter fine-tuning IndoBERT

Parameter	Nilai
Pre-trained model	indobenchmark/indobert-base-pl
Jumlah kelas	3
Max seq.length	128
Batch size	16
Learning rate	2×10^{-5}
Epoch maksimum	5
Optimizer	AdamW
Loss function	Weighted Cross-Entropy
Data split	80% / 10% / 10%

Evaluasi Kinerja Model

Evaluasi dilakukan menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Pada klasifikasi multikelas, macro average digunakan untuk memberikan bobot sama pada setiap kelas, sedangkan weighted average mempertimbangkan jumlah data pada setiap kelas.

$$Accuracy = \frac{\text{jumlah prediksi benar}}{\text{jumlah seluruh data}} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F_1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

Analisis Sentimen untuk Evaluasi Kualitas Produk

Distribusi sentimen digunakan sebagai informasi pendukung evaluasi kualitas produk berdasarkan persepsi pelanggan. Selain proporsi tiap kelas, penelitian menghitung Sentiment Quality Score (SQS), yaitu selisih jumlah ulasan positif dan negatif terhadap seluruh ulasan. SQS digunakan sebagai indikator persepsi pelanggan, bukan sebagai ukuran objektif kualitas fisik produk

$$SQS = \frac{N_{positif} - N_{negatif}}{N_{total}} \quad (12)$$

Tabel 6. Interpretasi Sentiment Quality Score

Rentang SQS	Interpretasi
0,60 – 1,00	Persepsi sangat positif
0,20 – <0,60	Cenderung positif
-0,20 – <0,20	Netral/campuran
-0,60 – <-0,20	Cenderung negatif
-1,00 – <-0,60	Sangat negatif

III. HASIL DAN PEMBAHASAN

Karakteristik Dataset

Setelah proses pembersihan dan eliminasi akhir terhadap teks kosong hasil preprocessing, dataset yang digunakan berjumlah 37.180 ulasan. Data dibagi menjadi 29.744 data pelatihan, 3.718 data validasi, dan 3.718 data pengujian. Distribusi kelas awal menunjukkan dominasi kelas positif, sehingga evaluasi model tidak cukup hanya menggunakan accuracy dan perlu mempertimbangkan macro F1-score.

Tabel 7. Pembagian dataset aktual

Dataset	Jumlah Data	Proporsi
Training	29.744	80%
Validation	3.718	10%
Testing	3.718	10%
Total	37.180	100%

ada data pengujian terdapat 3.452 data positif, 175 data netral, dan 91 data negatif. Komposisi ini menunjukkan ketidakseimbangan kelas yang sangat kuat dan berpotensi meningkatkan accuracy tanpa menjamin performa yang sama pada kelas minoritas.

Hasil Fine-Tuning IndoBERT

Class weighting menghasilkan bobot 13,6378 untuk kelas negatif, 7,0870 untuk kelas netral, dan 0,3590 untuk kelas positif. Bobot yang lebih tinggi pada kelas negatif dan netral dimaksudkan untuk mengurangi dominasi kelas positif selama pelatihan.

Tabel 8. Bobot kelas

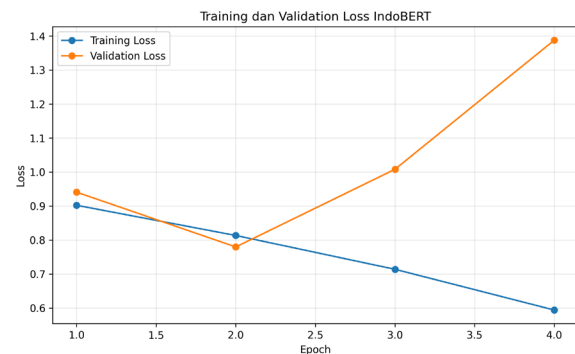
Kelas	Class Weight
Negatif	13,6378
Netral	7,0870
Positif	0,3590

Tabel 9. Perkembangan training dan validation

Epoch	Training Loss	Validation Loss	Val. Accuracy	Macro F1
1	0,9024	0,9410	93,17%	0,6185
2	0,8136	0,7799	92,33%	0,6236
3	0,7142	1,0083	92,76%	0,6142
4	0,5945	1,3877	93,36%	0,6187

Training loss terus menurun, tetapi validation loss meningkat setelah epoch kedua. Pola tersebut mengindikasikan kecenderungan overfitting. Model terbaik dipilih berdasarkan macro F1 pada data validasi dan pelatihan dihentikan melalui mekanisme

early stopping. Waktu pelatihan tercatat sekitar 1.041 detik atau 17,4 menit.



Gambar 2. Training loss dan validation loss IndoBERT

Evaluasi Kinerja Model IndoBERT

Pengujian pada 3.718 data menghasilkan accuracy 92,44%, weighted F1-score 92,67%, dan macro F1-score 63,36%. Perbedaan yang besar antara weighted dan macro F1 menunjukkan bahwa performa tinggi secara keseluruhan sangat dipengaruhi oleh kelas positif yang dominan.

Tabel 10. Kinerja keseluruhan model IndoBERT

Metrik	Nilai
Accuracy	0.9244
Macro Precision	0.6501
Macro Recall	0.6241
Macro F1-Score	0.6336
Weighted Precision	0.9297
Weighted Recall	0.9244
Weighted F1-Score	0.9267

Tabel 11. Classification report IndoBERT

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.6849	0.5495	0.6098	91
Netral	0.2972	0.3600	0.3256	175
Positif	0.9682	0.9629	0.9656	3452
macro avg	0.6501	0.6241	0.6336	3718
weighted avg	0.9297	0.9244	0.9267	3718

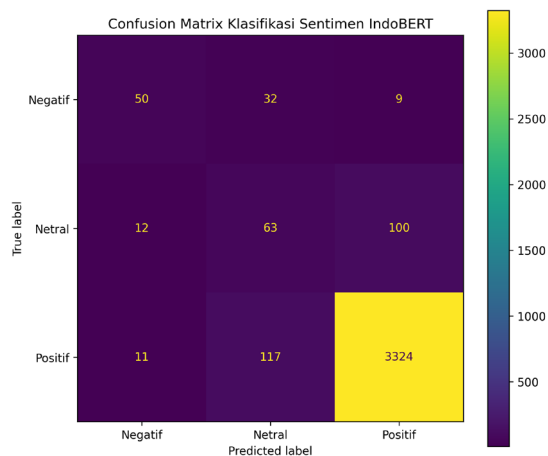
Kinerja terbaik diperoleh pada kelas positif dengan F1-score 96,56%. Kelas negatif memperoleh F1-score 60,98%, sedangkan kelas netral hanya 32,56%. Rendahnya performa kelas netral diduga berkaitan dengan jumlah data yang lebih sedikit serta ketidakselarasan antara rating tiga sebagai label

netral dan isi teks yang kadang masih memuat pujian atau keluhan kuat.

Analisis Confusion Matrix

Tabel 12. Classification report IndoBERT

Aktual / Prediksi	Negatif	Netral	Positif
Negatif	50	32	9
Netral	12	63	100
Positif	11	117	3.324



Gambar 3. Confusion matrix klasifikasi sentimen IndoBERT

Dari 91 ulasan negatif, 50 berhasil dikenali, 32 diprediksi netral, dan 9 diprediksi positif. Pada kelas netral, hanya 63 dari 175 ulasan yang dikenali dengan benar, sedangkan 100 ulasan bergeser menjadi positif. Sementara itu, 3.324 dari 3.452 ulasan positif diklasifikasikan secara tepat. Secara keseluruhan terdapat 281 kesalahan klasifikasi. Pola ini memperkuat bahwa kelas netral merupakan tantangan utama model.

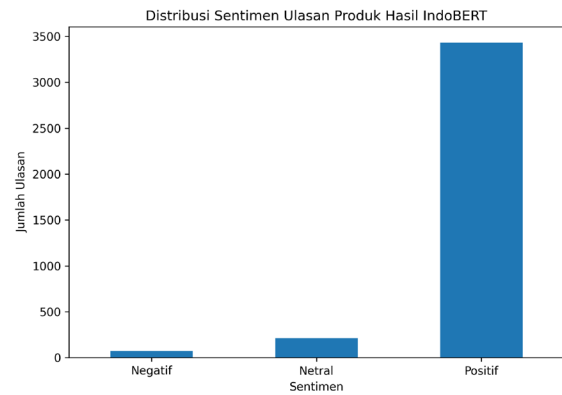
Distribusi Sentimen Hasil IndoBERT

Tabel 13. Classification report IndoBERT

Sentimen	Jumlah	Persentase
Positif	3.433	92,33%
Netral	212	5,70%
Negatif	73	1,96%
Total	3.718	100%

Dominasi sentimen positif menunjukkan persepsi pelanggan yang secara umum baik terhadap produk dalam dataset. Namun, temuan ini harus dibaca dengan mempertimbangkan bahwa dataset

awal memang sangat didominasi rating empat dan lima sehingga tidak dapat digeneralisasikan untuk seluruh produk Tokopedia.



Gambar 4. Distribusi sentimen hasil prediksi IndoBERT

Evaluasi Persepsi Kualitas Produk Menggunakan SQS

$$SQS = \frac{3.433 - 73}{3.718} = 0,9037$$

Tabel 14. Evaluasi persepsi kualitas produk

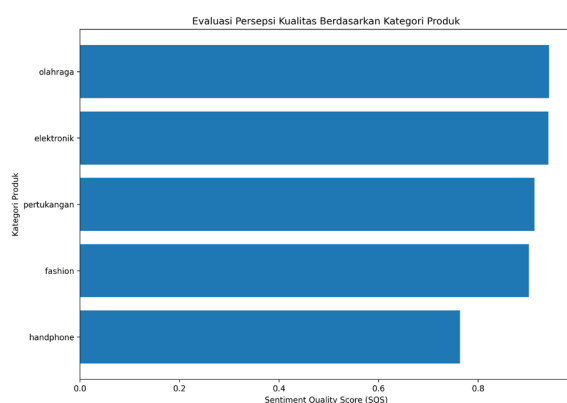
Indikator	Hasil
Jumlah Ulasan	3718
Sentimen Positif	3433
Sentimen Netral	212
Sentimen Negatif	73
Persentase Positif	92.33
Persentase Netral	5.7
Persentase Negatif	1.96
Sentiment Quality Score	0.9037
Evaluasi Persepsi Kualitas	Sangat Positif

Nilai SQS sebesar 0,9037 berada pada kategori sangat positif. Skor tersebut merepresentasikan kecenderungan persepsi pelanggan berdasarkan hasil klasifikasi sentimen, bukan pengukuran objektif kualitas fisik produk. Dengan demikian, SQS berfungsi sebagai indikator awal untuk mengidentifikasi produk atau kategori yang memerlukan peninjauan lebih lanjut.

Evaluasi Berdasarkan Kategori Produk

Tabel 15. Classification report IndoBERT

	Olahraga	Elektro nik	Pertuka ngan	Fashion	Hand phone
Ulasan	677	1460	173	829	579
Positif	640	1392	160	762	479
Netral	35	50	11	53	63
Negatif	2	18	2	14	37
Positif (%)	94.53%	95.34%	92.49%	91.92%	82.73%
Negatif (%)	0.30%	1.23%	1.16%	1.69%	6.39%
SQS	0.9424	0.9411	0.9133	0.9023	0.7634



Gambar 5. Sentiment Quality Score berdasarkan kategori produk

Kategori olahraga memperoleh SQS tertinggi sebesar 0,9424, diikuti elektronik sebesar 0,9411.

Kategori handphone memperoleh SQS terendah sebesar 0,7634 dan proporsi sentimen negatif tertinggi sebesar 6,39%. Analisis tingkat kategori memberikan informasi yang lebih spesifik karena SQS keseluruhan dapat menyembunyikan variasi pengalaman pelanggan pada kelompok produk tertentu.

Analisis Produk Berdasarkan SQS

Produk dengan jumlah ulasan yang sangat sedikit tidak langsung dibandingkan karena satu ulasan dapat menghasilkan SQS ekstrem. Oleh sebab itu, pembahasan produk difokuskan pada produk dengan minimal 20 ulasan agar interpretasi lebih stabil. Produk dengan SQS rendah dapat diprioritaskan untuk penelusuran isi ulasan negatif dan netral.

Analisis Ulasan Negatif

Kata seperti 'tidak', 'sesuai', 'kecewa', 'berfungsi', 'warna', 'gambar', dan 'salah' mengindikasikan bahwa sebagian sentimen negatif berkaitan dengan ketidaksesuaian produk, fungsi, kondisi barang, atau perbedaan dengan gambar/pesanan. Analisis ini dapat digunakan sebagai tahap awal untuk mengidentifikasi sumber ketidakpuasan, meskipun pendekatan frekuensi kata belum dapat membedakan secara eksplisit antara aspek kualitas fisik, pengiriman, pelayanan penjual, dan aspek lainnya.

Tabel 16. Contoh produk dengan SQS tinggi (minimal 20 ulasan)

Produk	Ulasan	Positif	Netral	Negatif	SQS
TP-LINK TL-WR 840N 300Mbps Wireless Router	23	23	0	0	1.0000
TINTA / CATRIDGE HP 678 BLACK / COLOR ORIGINAL 100%	80	79	1	0	0.9875
Switch HUB TP LINK 5 Port TL-SF1005D 10/100Mbps Network Lan Internet	75	74	1	0	0.9867
jam tangan pria Biden fashion analog bisnis pita jala - Biru	49	48	1	0	0.9796
TINTA / CATRIDGE HP 680 BLACK / COLOR ORIGINAL 100%	125	122	3	0	0.9760

Tabel 17. Contoh produk dengan SQS relatif terendah (minimal 20 ulasan)

Produk	Ulasan	Positif	Netral	Negatif	SQS
Tempelan / Perkat Gurita Hp Buat / untuk Gojek / Perkat Hp 3M Besar	22	15	6	1	0.6364
NOKIA 130 dual sim handphone hp	75	55	13	7	0.6400
Headset Bluetooth Mini S530 - Micro Sport Stereo Bluetooth Earphone	76	59	10	7	0.6842
Sepatu Sekolah Nike Roshe Run Full Hitam Anak Pria Wanita Cewe Cowo	26	21	3	2	0.7308
Mouse wireless / Bluetooth Airmouse Alcatroz by Powerlogic Resmi murah	45	38	3	4	0.7556

Tabel 18. Kata dominan pada ulasan negatif

Kata	Frekuensi
tidak	67
sudah	18
bisa	17
sesuai	14
dikirim	12
kecewa	10
berfungsi	9
padahal	9
baru	9
hari	9
warna	8
kirin	8
gambar	7
sangat	6
pake	6

Secara keseluruhan, IndoBERT menunjukkan kemampuan yang baik untuk klasifikasi sentimen ulasan produk berbahasa Indonesia dengan accuracy 92,44% dan weighted F1-score 92,67%. Namun, macro F1-score 63,36% memperlihatkan bahwa performa tersebut belum merata pada seluruh kelas. Kelas netral masih menjadi tantangan utama, yang berkaitan dengan ketidakseimbangan data dan karakter weak labeling berbasis rating.

Dari sisi tujuan penelitian, hasil klasifikasi berhasil dikembangkan menjadi informasi pendukung evaluasi persepsi kualitas produk. SQS keseluruhan sebesar 0,9037 menunjukkan kecenderungan sangat positif, sedangkan analisis per kategori memperlihatkan bahwa kategori handphone memiliki skor paling rendah dan proporsi sentimen negatif paling tinggi. Temuan ini menunjukkan nilai praktis analisis sentimen sebagai mekanisme penyaringan awal untuk menentukan kategori atau produk yang memerlukan evaluasi lebih lanjut.

Meskipun demikian, SQS tidak dapat diperlakukan sebagai pengukuran objektif kualitas fisik produk. Penelitian berikutnya dapat meningkatkan validitas dengan anotasi manual pada sebagian data, strategi penyeimbangan kelas yang lebih lanjut, serta aspect-based sentiment analysis untuk membedakan sentimen terhadap kualitas produk, kesesuaian deskripsi, harga, pengemasan, pengiriman, dan pelayanan penjual.

IV. KESIMPULAN

Berdasarkan hasil penelitian, model IndoBERT mampu melakukan klasifikasi sentimen ulasan pelanggan pada platform Tokopedia dengan kinerja yang baik. Pengujian terhadap 3.718 data menghasilkan accuracy sebesar 92,44%, weighted F1-score sebesar 92,67%, dan macro F1-score sebesar 63,36%. Kinerja terbaik diperoleh pada kelas positif dengan F1-score 96,56%, sedangkan kelas netral memperoleh F1-score 32,56%, yang menunjukkan pengaruh ketidakseimbangan kelas terhadap performa model.

Hasil klasifikasi digunakan untuk mendukung evaluasi persepsi pelanggan terhadap kualitas produk. Dari 3.718 data pengujian, 92,33% diklasifikasikan positif, 5,70% netral, dan 1,96% negatif. Sentiment Quality Score sebesar 0,9037 menunjukkan bahwa persepsi pelanggan secara keseluruhan berada pada kategori sangat positif. Kategori olahraga memperoleh SQS tertinggi sebesar 0,9424, sedangkan kategori handphone memperoleh SQS terendah sebesar 0,7634 dan proporsi sentimen negatif tertinggi sebesar 6,39%.

Secara keseluruhan, IndoBERT efektif untuk mengklasifikasikan sentimen ulasan produk berbahasa Indonesia dan hasil klasifikasinya dapat dimanfaatkan sebagai indikator awal untuk mendukung evaluasi kualitas produk berdasarkan persepsi pelanggan. Namun, SQS tidak merepresentasikan kualitas fisik produk secara objektif. Dominasi kelas positif dan penggunaan rating sebagai weak labeling menjadi keterbatasan penelitian. Penelitian selanjutnya disarankan melakukan anotasi manual pada sebagian data, menerapkan strategi penyeimbangan kelas yang lebih optimal, dan mengembangkan aspect-based sentiment analysis agar evaluasi dapat membedakan aspek kualitas produk, kesesuaian deskripsi, pengemasan, pengiriman, dan pelayanan penjual secara lebih spesifik.

V. REFERENSI

- Aji, B. P., & Aditya, C. S. K. (2025). Klasifikasi sentimen ulasan produk pada platform e-commerce di Indonesia dengan menggunakan model pre-trained IndoBERT. *Building of Informatics, Technology and Science (BITS)*,

- 6(4), 2491–2500.
- Alaiya, A., Nurdin, N., & Agusniar, C. (2025). Sentiment analysis of e-commerce product reviews on Tokopedia using Support Vector Machine. *Journal of Applied Informatics and Computing*, 9(5), 2869–2878.
- Aulia, M., & Hermawan, A. (2023). Analisis perbandingan algoritma SVM, Naïve Bayes, dan Perceptron untuk analisis sentimen ulasan produk Tokopedia. *Jurnal Media Informatika Budidarma*, 7(4), 1850–1859.
- Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment analysis on e-commerce product reviews using machine learning and deep learning algorithms: A bibliometric analysis, systematic literature review, challenges and future works. *International Journal of Information Management Data Insights*, 4(2), 100267.
- Dirfas, N. A., Rahmayanti, V., & Nastiti, S. (2024). Perbandingan kinerja pre-trained word embedding terhadap performa klasifikasi sentimen ulasan produk Tokopedia dengan Long Short-Term Memory (LSTM). *Building of Informatics, Technology and Science (BITS)*, 6(2), 878–889.
- Rasappan, P., Premkumar, M., Sinha, G., & Chandrasekaran, K. (2024). Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection. *Information Processing & Management*, 61.
- Ray, R. K., & Singh, A. (2025). From online reviews to smartwatch recommendation: An integrated aspect-based sentiment analysis framework. *Journal of Retailing and Consumer Services*, 82, 104059.
- Singh, U., Saraswat, A., Azad, H. K., Abhishek, K., & Shitharth, S. (2022). *Towards improving e-commerce customer review analysis for sentiment detection. Scientific Reports*, 12, 21983.
- Taneja, K., Vashishtha, J., & Ratnoo, S. (2024). Transformer based unsupervised learning approach for imbalanced text sentiment analysis of e-commerce reviews. *Procedia Computer Science*, 235, 2318–2331.