Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

ANALISIS KEPUASAN PELANGGAN PADA ULASAN PRODUK E-COMMERCE TOKOPEDIA MENGGUNAKAN BIDIRECTIONAL LONG SHORT-TERM MEMORY

Muhammad Faturrachman Iswan[1]*, Muhammad Faisal[2], Rizki Yusliana Bakti[3]
[1,2,3]Universitas Muhammadiyah Makassar, Indonesia
[1]105841103322@student.unismuh.ac.id, [2]muhfaisal@unismuh.ac.id,
[3]rizkiyusliana@unismuh.ac.id
* Coresponding Author

Info Artikel

Pengajuan : 01-08-2026
Peninjauan : 10-08-2026
Revisi : 26-08-2026
Diterima : 10-09-2026
Diterbitkan : 26-09-2026

Kata Kunci:
Bilstm; Analisis Sentimen; Kepuasan Pelanggan; Analisis Teks; E-Commerce.

Abstrak

Ulasan pelanggan pada platform e-commerce kerap menunjukkan ketidaksesuaian antara isi teks dan rating numerik yang diberikan, sehingga menyulitkan pelaku usaha dalam menilai kepuasan pelanggan secara objektif, terutama pada volume data yang besar. Penelitian ini bertujuan mengembangkan dan mengevaluasi model Bidirectional Long Short-Term Memory (BiLSTM) untuk mengklasifikasikan sentimen ulasan produk pada platform Tokopedia ke dalam kategori positif, netral, dan negatif sebagai representasi tingkat kepuasan pelanggan. Penelitian menggunakan pendekatan kuantitatif-eksperimental terhadap 4.060 data ulasan berbahasa Indonesia yang diperoleh melalui scraping. Data diproses melalui tahapan case folding, cleaning, tokenisasi, stopword removal, dan stemming, kemudian ditransformasikan menjadi representasi numerik melalui embedding layer berdimensi 128 sebelum dilatih menggunakan arsitektur BiLSTM dengan 64 unit dan optimizer Adam selama 10 epoch. Eksperimen dilakukan menggunakan dataset dengan distribusi kelas yang tidak seimbang (imbalanced dataset) tanpa penerapan teknik manipulasi data seperti SMOTE atau oversampling, maupun penyesuaian algoritma menggunakan class weight. Hasil pengujian pada 812 data uji dengan rasio 80:20 menunjukkan model mencapai akurasi 77,83%, dengan performa terbaik pada kelas positif (precision 0,84; recall 0,87; F1-score 0,86) dan performa terendah pada kelas netral dengan F1-score 0,44. Temuan ini menunjukkan bahwa BiLSTM mampu menangkap konteks dua arah dalam teks untuk klasifikasi sentimen berbahasa Indonesia, tetapi ketidakseimbangan distribusi kelas dan karakteristik kelas netral yang ambigu masih memengaruhi performa klasifikasi. Penelitian ini memberikan dasar bagi penerapan analisis sentimen untuk membantu pemantauan kepuasan pelanggan pada platform e-commerce secara otomatis

How to Cite : Iswan, M. F., Faisal, M., & Bakti, R. Y. (2026). Analisis kepuasan pelanggan pada ulasan produk e-commerce Tokopedia menggunakan bidirectional long short-term memory. Journal of Computer Science and Information Technology (JCSIT), 3(4), 531–541. https://doi.org/10.70248/jcsit.v3i4.4413

DOI: https://doi.org/10.70248/jcsit.v3i4.4413
This is an open access article under https://creativecommons.org/licenses/by/4.0/
Copyright© 2026 by Author. Published by Yayasan Nuraini Ibrahim Mandiri

PENDAHULUAN

Pertumbuhan transaksi digital telah menjadikan ulasan pelanggan sebagai sumber informasi strategis dalam ekosistem e-commerce karena mencerminkan pengalaman pengguna dan dapat memengaruhi keputusan pembelian. Namun, pemanfaatan ulasan tersebut tidak selalu sederhana. Teks naratif dapat menunjukkan sentimen yang berbeda dari rating bintang yang diberikan, sehingga analisis berbasis teks diperlukan untuk melengkapi informasi numerik (Huang, Zavareh, and Mustafa 2023). Persoalan ini semakin penting bagi pelaku usaha skala kecil dan menengah yang menerima

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

volume ulasan tinggi sehingga analisis manual menjadi kurang efisien dan konsisten.

Analisis sentimen berbasis deep learning dapat digunakan untuk mengklasifikasikan opini pelanggan ke dalam kategori positif, negatif, dan netral (Kamal and Astri 2025). Salah satu arsitektur yang relevan adalah Bidirectional Long Short-Term Memory (BiLSTM), yang memproses urutan teks dari arah maju dan mundur sehingga informasi konteks dari kedua arah dapat dipertimbangkan (Nasution et al. 2025).

Kajian terdahulu memperlihatkan beragam pendekatan analisis sentimen pada ulasan e-commerce Indonesia. (Azizah and Ernawati 2024) menggunakan K-Means Clustering pada 26.500 ulasan Tokopedia sehingga fokusnya adalah pengelompokan data tanpa evaluasi klasifikasi berlabel. Muzaki dkk. (Muzaki, Febriana, and Cholifah 2024) menerapkan Naive Bayes untuk klasifikasi dua kelas, yaitu positif dan negatif. Sementara itu, (Rafiah 2019) dan (Rahayu, N. K., & Sumadji 2019) membahas kepuasan pelanggan dan loyalitas pada konteks e-commerce, tetapi bukan klasifikasi teks tiga kelas. Kondisi tersebut menunjukkan ruang penelitian untuk mengevaluasi model deep learning bidirectional pada klasifikasi sentimen tiga kelas sekaligus melaporkan performa setiap kelas secara rinci.

Selain BiLSTM, perkembangan Natural Language Processing (NLP) menghadirkan Pre-trained Language Models (PLM), seperti IndoBERT, yang memanfaatkan representasi bahasa hasil pre-training. Model Transformer umumnya memiliki kapasitas representasi yang besar, tetapi kebutuhan sumber daya komputasinya juga dapat meningkat. Dalam konteks penelitian dengan dataset moderat dan kebutuhan komputasi yang terbatas, BiLSTM dapat digunakan sebagai pendekatan yang lebih sederhana untuk dipelajari dari data yang tersedia. Pernyataan ini tidak dimaksudkan sebagai klaim bahwa BiLSTM selalu lebih unggul daripada PLM; perbandingan langsung dengan model Transformer merupakan agenda penelitian lanjutan.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk: (1) mengembangkan model BiLSTM untuk klasifikasi sentimen ulasan produk e-commerce; (2) mengukur kinerja model dalam mengenali konteks kalimat ulasan; dan (3) menganalisis performa model pada masing-masing kelas sentimen sebagai representasi persepsi kepuasan pelanggan pada platform Tokopedia.

Ulasan Pelanggan dan Kepuasan pada E-Commerce

Ulasan pelanggan merupakan evaluasi tertulis pascatransaksi yang merefleksikan pengalaman, persepsi, dan tingkat kepuasan konsumen terhadap produk maupun layanan (Muhammad Zainal Abidin, Mochammad Alfan Rosid, and Ade Eviyanti 2025). Ulasan dapat menjadi sumber umpan balik bagi penjual untuk memahami pengalaman konsumen dan mengevaluasi aspek produk maupun layanan.

Analisis Sentimen dan Deep Learning

Analisis sentimen merupakan teknik komputasional untuk mengidentifikasi polaritas opini dalam teks dan mengelompokkannya, antara lain, menjadi positif, negatif, atau netral (Minaee et al. 2021). Pendekatan deep learning dapat mempelajari pola dan hubungan kontekstual dalam teks yang digunakan untuk klasifikasi (Sapanji, Hamdani, and Harahap 2023).

Recurrent Neural Network dan BiLSTM

Recurrent Neural Network (RNN) dirancang untuk memproses data sekuensial, tetapi dapat mengalami masalah vanishing gradient ketika mempelajari dependensi jangka panjang. Long Short-Term Memory (LSTM) mengatasi keterbatasan tersebut melalui mekanisme gate untuk mengatur aliran informasi. BiLSTM memperluas LSTM

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

dengan memproses sekuens dari arah forward dan backward sehingga representasi konteks menggabungkan informasi dari kedua arah (Algiffary and Sutabri 2023). Pendekatan tersebut relevan untuk teks ulasan yang maknanya dapat dipengaruhi oleh kata-kata pada posisi sebelum maupun sesudah suatu token. Studi terkait juga menunjukkan penggunaan CNN-BiLSTM untuk analisis sentimen berbasis aspek (Putro, Jacobus, and Kambey 2025).

Penelitian Terdahulu dan Kesenjangan Penelitian

Berikut merangkum posisi penelitian ini relatif terhadap studi-studi terdahulu.

Tabel 1. Perbandingan dengan Penelitian Terdahulu

Peneliti (Tahun)	Metode	Temuan Utama
Azizah & Ernawati (2024)	K-Means Clustering	Mengelompokkan 26.500 ulasan Tokopedia; mayoritas bersifat positif, namun tidak mengukur akurasi klasifikasi per kelas karena sifatnya unsupervised.
Muzaki dkk. (2024)	Naive Bayes	Membangun sistem klasifikasi sentimen dua kelas (positif-negatif) berbasis web; efektif untuk kasus biner namun tidak diuji pada skema tiga kelas dengan kelas netral.
Rafiah (2019)	Structural Equation Modeling (SEM)	Kepuasan memengaruhi loyalitas secara langsung maupun melalui kepercayaan sebagai mediator; bukan klasifikasi teks.
Nasution dkk. (2025)	BiLSTM vs GRU	BiLSTM digunakan untuk menangkap konteks dua arah pada ulasan e-commerce berbahasa Indonesia dibanding GRU.
Putro, Jacobus, & Kambey (2025)	CNN-BiLSTM + Attention	Kombinasi CNN-BiLSTM menunjukkan potensi pengembangan lanjutan untuk analisis sentimen berbasis aspek.

Merujuk pada Tabel 1, penelitian terdahulu menggunakan pendekatan yang beragam, mulai dari clustering, Naive Bayes, hingga model deep learning. Penelitian ini memusatkan evaluasi pada BiLSTM untuk klasifikasi sentimen tiga kelas (positif, netral, negatif) pada ulasan Tokopedia dan melaporkan performa per kelas. Fokus tersebut penting karena distribusi dataset tidak seimbang dan kelas netral menunjukkan tingkat kesalahan yang relatif tinggi pada hasil pengujian.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif-eksperimental, yaitu membangun dan menguji model klasifikasi berbasis deep learning terhadap data ulasan pelanggan, kemudian mengevaluasi performanya menggunakan metrik statistik standar.

Sumber dan Teknik Pengumpulan Data

Data penelitian berupa 4.060 ulasan produk berbahasa Indonesia dari platform Tokopedia yang dikumpulkan melalui proses scraping. Setiap data memuat atribut Review, Rating, dan Sentiment. Review berisi teks ulasan, Rating berisi penilaian numerik pelanggan, sedangkan Sentiment digunakan sebagai target klasifikasi. Karena dokumen sumber yang tersedia tidak menjelaskan secara eksplisit prosedur pembentukan label Sentiment pada seluruh 4.060 data, artikel ini tidak mengklaim skema tersebut sebagai ground truth dataset asli. Sebagai pelengkap revisi, Tabel 2 menyajikan contoh data dummy dan aturan pelabelan ilustratif berdasarkan rating: rating 1–2 dikategorikan negatif, rating 3 netral, dan rating 4–5 positif. Aturan ilustratif ini digunakan hanya untuk memperjelas mekanisme ground truth dan harus diganti

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

dengan prosedur anotasi sebenarnya apabila tersedia.

Tabel 2. Contoh Data Dummy dan Aturan Ground Truth Ilustratif

No	Ulasan Dummy	Rating	Label Ground Truth
1	Barangnya bagus, sesuai dengan deskripsi dan kualitasnya sangat baik.	5	Positif
2	Produk sudah sampai dengan aman, packing rapi dan barang tidak rusak.	5	Positif
3	Barang sudah diterima sesuai pesanan.	4	Positif
4	Produk sesuai dengan deskripsi yang tertera di toko.	3	Netral
5	Barang sudah sampai dan dapat digunakan dengan baik.	3	Netral
6	Pengiriman membutuhkan waktu sekitar tiga hari dan barang sudah diterima.	3	Netral
7	Barang yang diterima rusak dan tidak dapat digunakan.	1	Negatif
8	Kualitas produk sangat mengecewakan dan tidak sesuai dengan deskripsi.	1	Negatif
9	Barang berbeda dengan yang saya pesan dan kualitasnya kurang baik.	2	Negatif

Teknik Analisis Data

Implementasi yang terdokumentasi melakukan preprocessing melalui case folding, cleaning, stopword removal, tokenisasi, dan stemming sebagaimana dijelaskan dalam naskah skripsi. Namun, source code preprocessing yang tersedia secara eksplisit menunjukkan fungsi clean_text dan remove_stopwords, sedangkan normalisasi slang dan penanganan negasi belum terdokumentasi. Karena kata seperti “tidak”, “bukan”, “belum”, dan “kurang” dapat membawa informasi polaritas, daftar stopword perlu diperiksa agar kata negasi tidak dihapus. Untuk menjawab masukan reviewer mengenai slang, normalisasi kata tidak baku dan singkatan perlu ditambahkan ke implementasi kemudian diuji ulang. Dampak stemming juga perlu diuji melalui ablation study sebelum dinyatakan sebagai bagian dari konfigurasi final.

Eksperimen yang terdokumentasi menggunakan train_test_split dengan rasio 80:20 dan stratifikasi label, menghasilkan 3.248 data latih dan 812 data uji. Berdasarkan source code skripsi, model dilatih menggunakan class_weight dan EarlyStopping, dengan data uji juga diberikan sebagai validation_data. Untuk eksperimen final, data uji sebaiknya dipisahkan dan tidak digunakan sebagai validation_data agar evaluasi tetap independen. Selain itu, hasil Stratified K-Fold Cross-Validation 5-fold belum terdokumentasi pada sumber yang tersedia dan harus dihitung dari dataset asli sebelum dilaporkan sebagai hasil final. Jika reviewer mensyaratkan baseline tanpa class weighting, eksperimen tersebut juga harus dijalankan ulang dan dilaporkan terpisah dari konfigurasi source code yang menggunakan class_weight.

Distribusi kelas yang tidak seimbang perlu dipertahankan atau ditangani secara eksplisit sesuai tujuan eksperimen. Confusion matrix yang tersedia berasal dari eksperimen holdout dan menunjukkan bahwa kesalahan cukup banyak terjadi pada perpindahan kelas netral ke positif. Untuk eksperimen tambahan, Stratified K-Fold Cross-Validation dapat digunakan untuk mempertahankan proporsi kelas pada setiap fold. Teknik resampling, class weighting, atau Focal Loss sebaiknya diperlakukan sebagai eksperimen terpisah agar pengaruh masing-masing metode dapat diukur secara jelas.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

Tabel 3. Parameter Arsitektur dan Pelatihan Model BiLSTM

Parameter	Nilai
Ukuran kosakata maksimum (max words)	5.000 kata
Panjang urutan maksimum (max length)	100 token
Dimensi embedding	128
Unit BiLSTM	64 (per arah)
Dropout	0,5 dan 0,3
Neuron dense layer	64 (aktivasi ReLU)
Fungsi aktivasi output	Softmax (3 kelas)
Optimizer	Adam
Fungsi loss	Categorical crossentropy
Jumlah epoch	15 (maksimum; EarlyStopping digunakan)
Batch size	32
Rasio data latih : data uji	80% : 20% (3.248 : 812)
Total parameter model	747.267
Class weight	Balanced (digunakan pada source code skripsi)
Early stopping	monitor val_loss, patience 3, restore_best_weights=True

Kinerja model dievaluasi menggunakan confusion matrix sebagai dasar penghitungan accuracy, precision, recall, dan F1-score(Powers 2020). Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data uji, precision menunjukkan ketepatan prediksi suatu kelas, recall menunjukkan kemampuan menemukan data yang benar-benar termasuk dalam suatu kelas, sedangkan F1-score merupakan rata-rata harmonis precision dan recall. Pada dataset tidak seimbang, Macro F1 digunakan untuk melihat performa rata-rata antarkelas tanpa memberi bobot berdasarkan jumlah sampel.

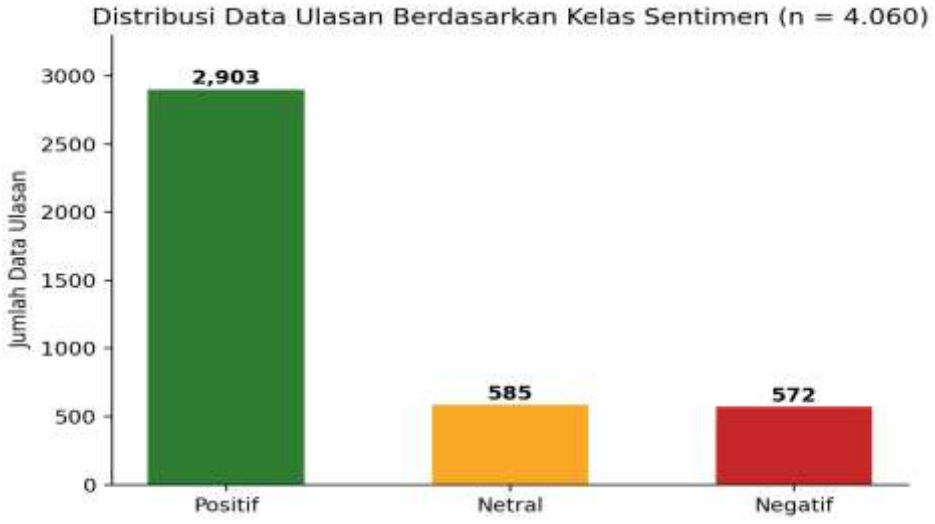
HASIL PENELITIAN DAN PEMBAHASAN

Karakteristik Dataset

Dari 4.060 data ulasan yang terkumpul, distribusi kelas sentimen menunjukkan dominasi kelas positif. Dataset terdiri atas 2.903 data positif, 585 data netral, dan 572 data negatif.

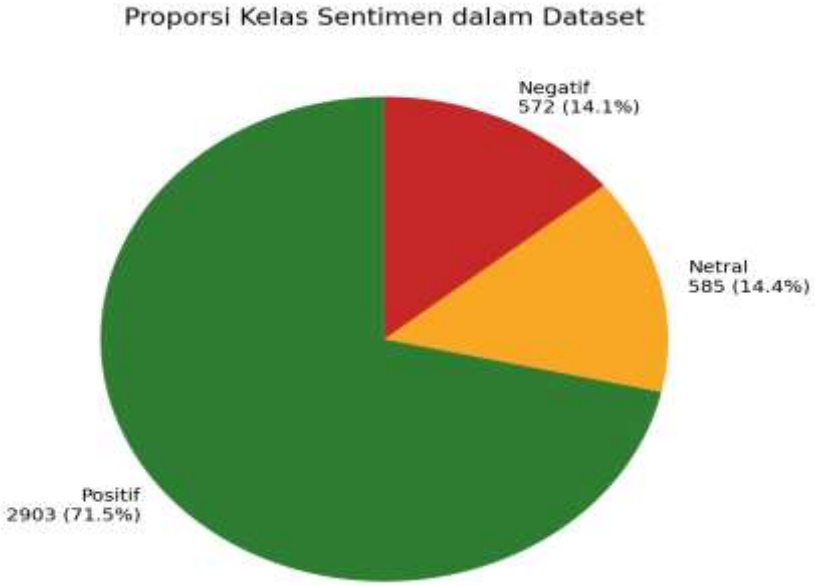
Tabel 4. Distribusi Dataset Berdasarkan Kelas Sentimen

Kelas Sentimen	Jumlah Data	Contoh Ulasan
Positif	2.903	Sangat baik
Netral	585	Cukup
Negatif	572	Sangat buruk
Total	4.060	-



Gambar 1. Distribusi Data Ulasan Berdasarkan Kelas Sentimen

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

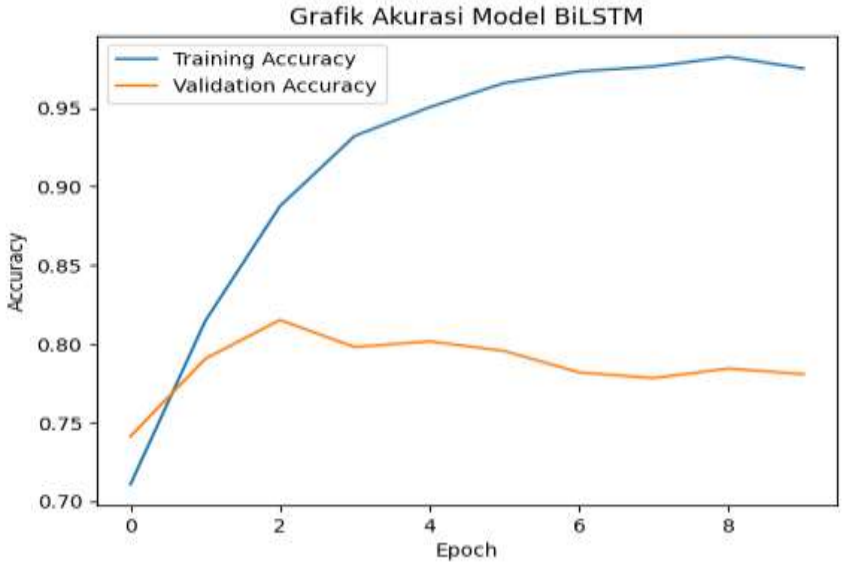


Gambar 2. Proporsi Kelas Sentimen dalam Dataset

Kelas positif memiliki proporsi 71,50%, sedangkan kelas netral 14,41% dan negatif 14,09%. Ketimpangan tersebut menunjukkan bahwa jumlah contoh untuk kelas positif jauh lebih besar dibandingkan dua kelas lainnya. Kondisi ini perlu diperhatikan karena model dapat memperoleh lebih banyak contoh pembelajaran untuk kelas mayoritas.

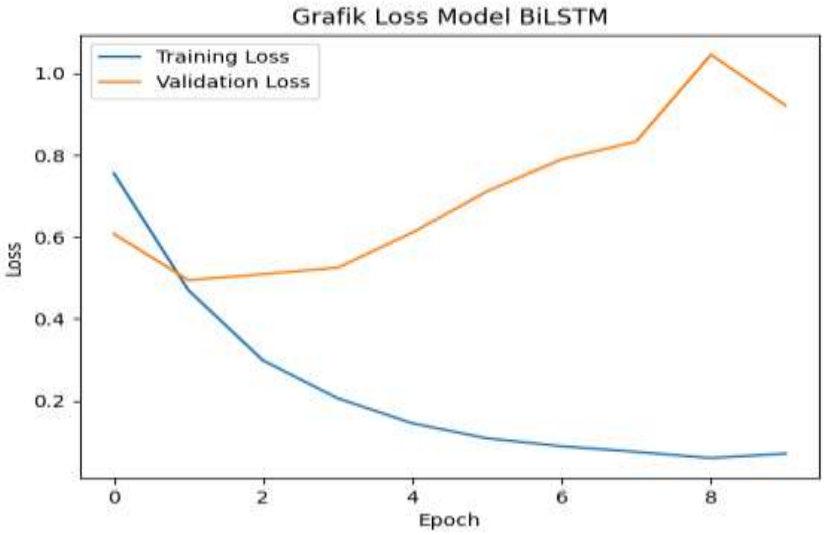
Proses Pelatihan Model

Selama proses pelatihan, akurasi dan loss pada data training dan validation dicatat pada setiap epoch. Grafik hasil pelatihan pada skripsi menunjukkan bahwa training accuracy meningkat hingga sekitar 0,97, sedangkan validation accuracy berada sekitar 0,78 pada akhir pelatihan. Pada saat yang sama, training loss terus menurun, sementara validation loss meningkat setelah beberapa epoch. Pola divergensi tersebut merupakan indikasi overfitting, sehingga hasil pelatihan tidak dapat dinyatakan sebagai bukti bahwa overfitting tidak terjadi.

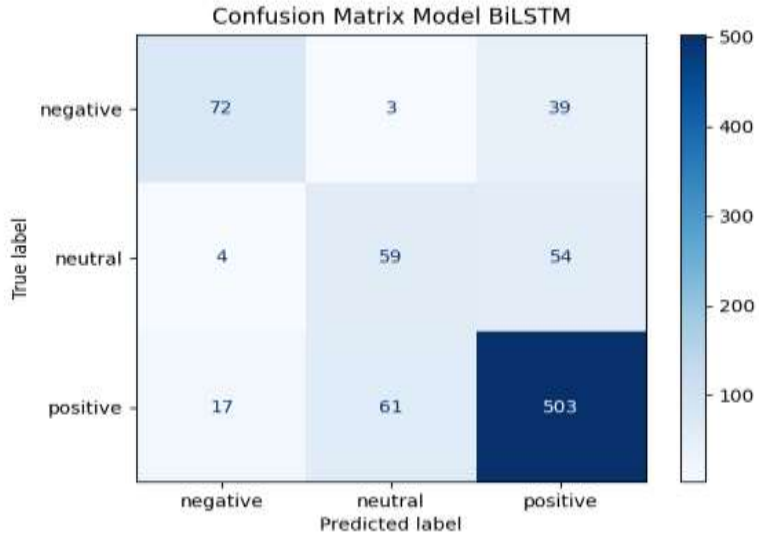


Gambar 3. Grafik Accuracy Training dan Validation per Epoch

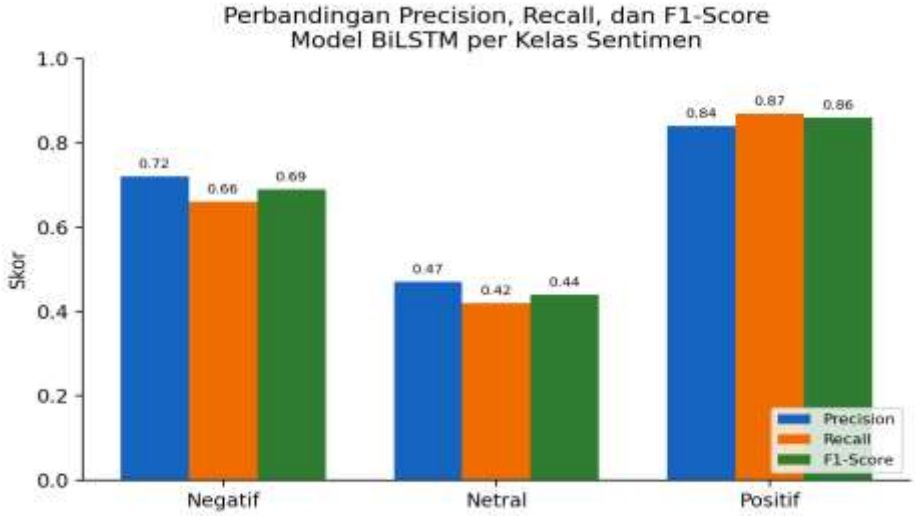
Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541



Gambar 4. Grafik Loss Training dan Validation per Epoch



Gambar 5. Confusion Matrix Model BiLSTM



Gambar 6. Perbandingan Precision, Recall, dan F1-Score per Kelas Sentimen

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

5. Confusion Matrix Numerik Model BiLSTM

Label Aktual / Prediksi	Negatif	Netral	Positif	Total
Negatif	72	3	39	114
Netral	4	59	54	117
Positif	17	61	503	581
Total	93	123	596	812

Evaluasi Kinerja Model

Pengujian pada 812 data uji menghasilkan akurasi sebesar 78,08% berdasarkan confusion matrix yang tersedia. Matriks tersebut berisi 634 prediksi benar dari 812 data. Perhitungan metrik dari matriks menghasilkan Macro F1-score sekitar 0,68.

Confusion matrix pada Gambar 5 menunjukkan jumlah prediksi aktual dan prediksi model untuk tiga kelas. Pada kelas negatif, 72 data diprediksi benar sebagai negatif, 3 diprediksi netral, dan 39 diprediksi positif. Pada kelas netral, 59 data diprediksi benar sebagai netral, 4 diprediksi negatif, dan 54 diprediksi positif. Pada kelas positif, 503 data diprediksi benar sebagai positif, 17 diprediksi negatif, dan 61 diprediksi netral. Pola tersebut memperlihatkan bahwa kesalahan terbesar untuk kelas netral adalah ketika data netral diprediksi sebagai positif.

Tabel 6. Hasil Classification Report yang Dihitung dari Confusion Matrix

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,77	0,63	0,70	114
Netral	0,48	0,50	0,49	117
Positif	0,84	0,87	0,85	581
Akurasi	-	-	0,78	812
Rata-rata makro	0,70	0,67	0,68	812
Rata-rata tertimbang	0,78	0,78	0,78	812

Berdasarkan confusion matrix, kelas positif memiliki precision sekitar 0,84, recall sekitar 0,87, dan F1-score sekitar 0,85. Kelas negatif memiliki precision sekitar 0,77, recall sekitar 0,63, dan F1-score sekitar 0,70. Kelas netral memiliki precision sekitar 0,48, recall sekitar 0,50, dan F1-score sekitar 0,49. Dengan demikian, kelas netral merupakan kelas yang paling sulit diklasifikasikan secara konsisten. Perbedaan ini juga menunjukkan bahwa akurasi keseluruhan perlu dibaca bersama metrik per kelas, terutama pada dataset dengan distribusi yang tidak seimbang.

Analisis kesalahan berdasarkan confusion matrix menunjukkan bahwa 54 data aktual netral diprediksi sebagai positif dan 4 data aktual netral diprediksi sebagai negatif. Untuk memenuhi kebutuhan interpretasi error analysis, Tabel 7 menyajikan contoh ulasan dummy yang menggambarkan pola kesalahan tersebut. Contoh ini bersifat ilustratif dan bukan klaim bahwa kalimat yang ditampilkan merupakan baris asli dari 812 data uji. Pola kesalahan yang ditunjukkan terutama berkaitan dengan kata evaluatif ambigu, negasi, kata penguat, dan campuran opini.

Tabel 7. Contoh Error Analysis Menggunakan Data Dummy

Ulasan Dummy	Label Aktual	Prediksi BiLSTM	Interpretasi
Barang sudah sampai, kualitasnya lumayan dan bisa digunakan.	Netral	Positif	Kata "lumayan" dan "bisa digunakan" dapat memberi sinyal positif.
Produknya biasa saja, sesuai dengan harga yang diberikan.	Netral	Positif	Kata "sesuai" dapat lebih dominan ditangkap sebagai indikator positif.
Barangnya tidak terlalu	Netral	Negatif	Kata negasi "tidak" dan

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 531-541

Ulasan Dummy	Label Aktual	Prediksi BiLSTM	Interpretasi
bagus, tetapi masih bisa digunakan.			"bagus" dapat menghasilkan pola yang ambigu.
Kualitasnya kurang, namun barang masih berfungsi.	Netral	Negatif	Kata "kurang" dapat menjadi indikator negatif meskipun terdapat informasi positif.
Pengiriman agak lama tetapi barangnya sesuai pesanan.	Netral	Positif	Model dapat lebih menekankan frasa "sesuai pesanan".

Weighted F1-score yang dihitung dari confusion matrix sekitar 0,78, sedangkan Macro F1-score sekitar 0,68. Perbedaan tersebut menunjukkan bahwa metrik berbobot banyak dipengaruhi oleh kelas positif yang memiliki jumlah data lebih besar. Oleh karena itu, Macro F1 lebih informatif untuk melihat kesenjangan performa antarkelas pada dataset ini.

Perbandingan dengan Penelitian Terdahulu dan Kontribusi Penelitian

Dibandingkan penelitian Naive Bayes yang dikaji pada dua kelas (Wu, Xia, and Tian 2025), penelitian ini menggunakan tiga kelas sehingga evaluasi mencakup sentimen netral. BiLSTM memproses urutan teks dari dua arah dan dapat digunakan untuk mempelajari konteks sekuensial. Namun, hasil pengujian menunjukkan bahwa kemampuan tersebut belum menghilangkan kesalahan klasifikasi pada kelas netral. Temuan ini juga perlu dibaca dalam konteks konfigurasi eksperimen yang tersedia, yaitu holdout 80:20 dan implementasi source code yang menggunakan class weighting.

Kontribusi penelitian ini meliputi evaluasi BiLSTM pada klasifikasi sentimen tiga kelas untuk ulasan Tokopedia, pelaporan performa per kelas, serta identifikasi pola kesalahan yang dominan pada kelas netral. Hasil confusion matrix memberikan informasi bahwa kesalahan netral paling banyak berpindah ke kelas positif. Untuk pengembangan berikutnya, arsitektur dapat diperluas melalui attention, perubahan dimensi embedding, pretrained embedding seperti FastText atau model bahasa pralatih, serta strategi penanganan imbalance seperti resampling, class weighting, atau Focal Loss.

Dalam implementasi praktis, keluaran model dapat diintegrasikan ke dashboard analytics yang menampilkan proporsi sentimen berdasarkan periode, produk, dan kategori serta daftar ulasan yang memerlukan perhatian. Integrasi dashboard tersebut merupakan rancangan pemanfaatan hasil penelitian dan belum diuji sebagai bagian dari sistem pada penelitian ini.

KESIMPULAN

Penelitian ini mengembangkan model Bidirectional Long Short-Term Memory (BiLSTM) untuk klasifikasi sentimen pada 4.060 ulasan produk Tokopedia. Pada eksperimen holdout 80:20 yang terdokumentasi, confusion matrix menghasilkan 634 prediksi benar dari 812 data atau akurasi sekitar 78,08%. Berdasarkan matriks tersebut, Macro F1-score sekitar 0,68, dengan F1-score sekitar 0,70 pada kelas negatif, 0,49 pada kelas netral, dan 0,85 pada kelas positif. Hasil ini menunjukkan adanya kesenjangan performa antarkelas, terutama pada kelas netral. Grafik pelatihan juga menunjukkan indikasi overfitting karena training loss terus menurun sementara validation loss meningkat setelah beberapa epoch. Artikel ini menambahkan contoh

data dummy untuk menjelaskan mekanisme ground truth dan error analysis karena prosedur pelabelan asli serta pasangan teks–prediksi dari data uji tidak terdokumentasi lengkap pada dokumen sumber. Dengan demikian, contoh dummy tersebut tidak digunakan sebagai bukti tambahan kinerja model. Keterbatasan penelitian mencakup belum tersedianya hasil Stratified K-Fold Cross-Validation, belum terdokumentasinya secara eksplisit prosedur pembentukan ground truth pada seluruh dataset, serta variasi domain produk yang belum diuji secara rinci. Penelitian selanjutnya disarankan melakukan Stratified K-Fold, menguji normalisasi slang dan penanganan negasi, membandingkan pretrained embedding seperti FastText atau model bahasa pralatih, serta menguji resampling, class weighting, Focal Loss, perubahan dimensi embedding, dan mekanisme attention secara terkontrol.

Pernyataan mengenai Penggunaan Kecerdasan Buatan (AI)

AI digunakan dalam penyusunan naskah ini. Penulis menggunakan perangkat kecerdasan buatan semata-mata untuk membantu perbaikan bahasa, pemeriksaan tata bahasa, serta penyuntingan demi kejelasan dan keterbacaan. Penulis tetap bertanggung jawab penuh atas akurasi, orisinalitas, integritas, dan konten akademis naskah tersebut. Seluruh konten yang dibantu oleh AI telah ditinjau dan diverifikasi oleh penulis sebelum naskah diserahkan.

Konflik kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

Kontribusi penulis

Konseptualisasi, Muhammad Faturrachman Iswan, Muhammad Faisal, dan Rizki Yusliana Bakti; pengumpulan data, Muhammad Faturrachman Iswan; validasi data, Muhammad Faisal dan Rizki Yusliana Bakti; analisis formal, Muhammad Faturrachman Iswan; penulisan—penyusunan draf awal, Muhammad Faturrachman Iswan; penulisan—peninjauan dan penyuntingan, Muhammad Faisal dan Rizki Yusliana Bakti; visualisasi, Muhammad Faturrachman Iswan. Seluruh penulis telah membaca dan menyetujui versi akhir naskah ini.

Ucapan terima kasih

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Makassar, Fakultas Teknik, Program Studi Informatika, serta pihak-pihak yang telah memberikan dukungan dan fasilitas selama pelaksanaan penelitian dan penyusunan artikel ini.

DAFTAR PUSTAKA

Algiffary, Arief, and Tata Sutabri. 2023. “Indonesian Journal of Computer Science.” *Indonesian Journal of Computer Science* 12(2): 284–301. http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135.

Azizah, Nur Aviani, and Tati Ernawati. 2024. “Analisis Sentimen Ulasan Pelanggan Terhadap Produk Roti Macan Artisan Di E-Commerce Tokopedia Menggunakan Metode Clustering.” *Journal of Information System, Applied, Management, Accounting and Research* 8(3): 580. doi:10.52362/jisamar.v8i3.1576.

Huang, Huang, Adeleh Asemi Zavareh, and Mumtaz Begum Mustafa. 2023. “Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions.” *IEEE Access* 11: 90367–82. doi:10.1109/ACCESS.2023.3307308.

Kamal, Ahmad, and Renita Astri. 2025. “Eksplorasi Sentimen Pengguna Pada Aplikasi E-

Commerce Dengan Deep Learning." *Jurnal Teknologi Dan Sistem Informasi Bisnis* 7(3): 435–41. doi:10.47233/jteksis.v7i3.2010.

Minaee, Shervin, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. "Deep Learning Based Text Classification: A Comprehensive Review." 1(1): 1–43. http://arxiv.org/abs/2004.03705.

Muhammad Zainal Abidin, Mochammad Alfan Rosid, and Ade Eviyanti. 2025. "Analisis Sentimen Pengguna Aplikasi Shopee Pada Google Play Store Menggunakan Metode Random Forest." *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer* 4(3): 265–76. doi:10.55123/storage.v4i3.6027.

Muzaki, Assami, Vici Febriana, and Wahyu Nur Cholifah. 2024. "Analisis Sentimen Pada Ulasan Produk Di E-Commerce Dengan Metode Naive Bayes." *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)* 5(4): 758–65. doi:10.30998/jrami.v5i4.9647.

Nasution, Khairunnisa, Khairun Saddami, Roslidar Roslidar, Akhyar Akhyar, Fathurrahman Fathurrahman, and Niza Aulia. 2025. "Comparative Study of BiLSTM and GRU for Sentiment Analysis on Indonesian E-Commerce Product Reviews Using Deep Sequential Modeling." *Jurnal Teknik Informatika (Jutif)* 6(4): 1881–96. doi:10.52436/1.jutif.2025.6.4.4878.

Powers, David M. W. 2020. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation." : 37–63. http://arxiv.org/abs/2010.16061.

Putro, Obedient Boas Jr, Agustinus Jacobus, and Feisy Diane Kambey. 2025. "Aspect-Based Sentiment Analysis Product Review CNN and Bidirectional LSTM." *Jurnal Teknik Informatika* 20(2): 117–24.

Rafiah, Kurnia Khafidhatur. 2019. "Analisis Pengaruh Kepuasan Pelanggan Dan Kepercayaan Pelanggan Terhadap Loyalitas Pelanggan Dalam Berbelanja Melalui E-Commerce Di Indonesia." *Al Tijarah* 5(1): 46. doi:10.21111/tijarah.v5i1.3621.

Rahayu, N. K., & Sumadji, S. 2019. "Analisis Kepuasan Pelanggan Pada Layanan E-Commerce." *Jurnal Manajemen dan Pemasaran Jasa* 3(1): 1–12.

Sapanji, R. A. E. Virgana Targa, Dani Hamdani, and Parlindungan Harahap. 2023. "Sentiment Analysis of the Top 5 E-Commerce Platforms in Indonesia Using Text Mining and Natural Language Processing (NLP)." *Journal of Applied Informatics and Computing* 7(2): 202–11. doi:10.30871/jaic.v7i2.6517.

Wu, Qianye, Chengxuan Xia, and Sixuan Tian. 2025. "AI-Driven Sentiment Analytics: Unlocking Business Value in the E-Commerce Landscape." http://arxiv.org/abs/2504.08738.