Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

Analisis Sentimen Terhadap Ulasan Google Play Store Aplikasi Lazada, Shopee, dan Tokopedia Menggunakan Algoritma IndoBERT

Afra Rihadatul Aisy*, Giat Karyono

Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Amikom Purwokerto, Banyumas, Indonesia
Email: [1,*]afraaisy11@gmail.com, [2]giantmercy123@gmail.com
Email Penulis Korespondensi: afraaisy11@gmail.com
Submitted: **18/11/2025**; Accepted: **30/12/2025**; Published**: 31/12/2025**

Abstrak−Pertumbuhan e-commerce menghasilkan banyak ulasan pengguna yang menjadi sumber penting untuk memahami kepuasan dan persepsi konsumen. Namun, analisis manual terhadap ulasan yang tidak terstruktur dan menggunakan bahasa informal kurang efektif. Selain itu, pendekatan sentimen konvensional sering tidak mampu menangkap variasi linguistik Bahasa Indonesia. Penelitian ini menggunakan model bahasa kontekstual IndoBERT untuk mengklasifikasikan sentimen ulasan aplikasi e-commerce pada Shopee, Tokopedia, dan Lazada. Data dikumpulkan melalui web scraping berjumlah 12.000 data dengan masing-masing aplikasi berjumlah 4.000, diberi label berdasarkan rating, diproses melalui tahapan preprocessing, diseimbangkan menggunakan Random Oversampling, dan dilatih untuk klasifikasi tiga kelas sentimen. Evaluasi menunjukkan Macro F1-Score 0,90, menandakan kinerja kuat pada seluruh kelas sentimen, termasuk kelas minoritas. Hasil ini menegaskan efektivitas IndoBERT dengan penanganan ketidakseimbangan data dalam analisis sentimen berbahasa Indonesia.

Kata Kunci: Analisis Sentimen; IndoBERT; E-Commerce; Natural Language Processing; Fine-Tuning; Oversampling

Abstract−The growth of e-commerce has generated many user reviews, which are an important source for understanding consumer satisfaction and perceptions. However, manual analysis of unstructured reviews that use informal language is ineffective. In addition, conventional sentiment analysis approaches are often unable to capture the linguistic variations of the Indonesian language. This study uses the IndoBERT contextual language model to classify the sentiment of e-commerce application reviews on Shopee, Tokopedia, and Lazada. Data was collected through web scraping, amounting to 12,000 data points, with 4,000 for each application, labeled based on ratings, processed through preprocessing stages, balanced using Random Oversampling, and trained for three-class sentiment classification. The evaluation showed an Macro F1-Score of 0.90, indicating strong performance across all sentiment classes, including minority classes. These results confirm the effectiveness of IndoBERT in handling data imbalance in Indonesian sentiment analysis.

Keywords: Sentiment Analysis; IndoBERT; E-Commerce; Natural Language Processing; Fine-Tuning; Oversampling

1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah secara mendasar bagaimana masyarakat melakukan transaksi, khususnya dalam aktivitas jual beli. Saat ini, konsumen tidak perlu lagi datang ke toko secara langsung, hanya dengan perangkat digital, mereka dapat berbelanja kapan pun dan di mana pun [1][2]. Perubahan ini mendorong tumbuhnya platform e-commerce sebagai solusi yang menjawab kebutuhan akan kemudahan, kecepatan, dan aksesibilitas [3][4].

Meningkatnya penggunaan e-commerce menghasilkan jumlah besar ulasan pengguna berupa teks bebas yang menggambarkan pengalaman, tingkat kepuasan, hingga keluhan mereka terhadap layanan aplikasi. Informasi tersebut tidak hanya bermanfaat sebagai referensi bagi calon konsumen, tetapi juga sebagai sumber umpan balik langsung bagi penyedia layanan untuk memahami persepsi pengguna [1][4][5]. Akan tetapi, besarnya volume ulasan membuat proses analisis secara manual menjadi tidak efisien. Pada titik ini, analisis sentimen dalam bidang Natural Language Processing (NLP) menjadi relevan, karena teknik tersebut dapat mengelompokkan opini pengguna secara otomatis ke dalam sentimen positif, negatif, atau netral [2] [6].

Berbagai penelitian sebelumnya telah menggunakan algoritma machine learning tradisional seperti Naïve Bayes, SVM, KNN, dan Random Forest untuk analisis sentimen ulasan e-commerce di Indonesia. Penelitian [3] melaporkan akurasi sebesar 85% menggunakan Naïve Bayes pada dataset yang relatif kecil (2.000 ulasan), dengan evaluasi yang didominasi oleh metrik akurasi. Sementara itu, penelitian [7] melaporkan akurasi hingga 92% pada dataset yang lebih besar. Namun, analisis lebih lanjut melalui classification report menunjukkan adanya ketimpangan performa antar kelas, di mana kelas minoritas memiliki nilai precision dan F1-score yang jauh lebih rendah dibandingkan kelas mayoritas. Hal ini tercermin dari nilai macro F1-score yang hanya mencapai 0.79, meskipun nilai akurasi terlihat tinggi. Temuan ini mengindikasikan bahwa penggunaan akurasi sebagai metrik utama dapat menutupi kelemahan model dalam menangani data yang tidak seimbang, serta menunjukkan keterbatasan pendekatan berbasis TF-IDF dan Naïve Bayes dalam menangkap konteks dan nuansa bahasa ulasan e-commerce Indonesia. Sementara itu, [8] menggunakan Random Forest dengan pelabelan otomatis berbasis leksikon VADER yang awalnya dirancang untuk bahasa Inggris dan mengakui bahwa proses terjemahan berpotensi menyebabkan kehilangan makna, sehingga menjadi batasan utama dalam konteks bahasa Indonesia. Penelitian oleh [6] yang membandingkan TF-IDF dan Count Vectorizer, juga menunjukkan bahwa akurasi model masih berkisar di angka 80%, mengindikasikan adanya ruang signifikan untuk perbaikan. Secara umum, metode tradisional ini kesulitan menangani ciri khas bahasa alami pengguna Indonesia, seperti slang, ekspresi informal, ironi, atau konteks budaya yang tidak literal [5].

Keterbatasan tersebut menunjukkan perlunya pendekatan yang lebih adaptif terhadap kompleksitas linguistik bahasa Indonesia. IndoBERT sebuah model pre-trained berbasis arsitektur model transformer yang dikembangkan

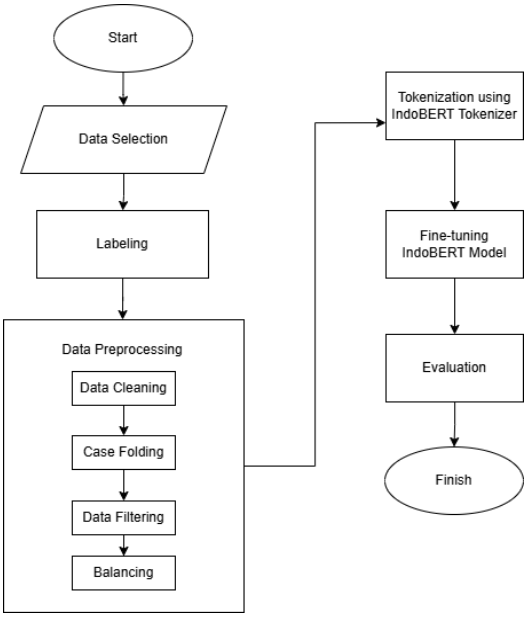
Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

menggunakan korpus bahasa Indonesia dalam skala besar mampu menangkap konteks dengan lebih mendalam dibandingkan pendekatan statistik tradisional [7]. Berbeda dengan TF-IDF yang memperlakukan kata secara terpisah, IndoBERT memahami makna sebuah kata berdasarkan posisinya dalam kalimat, sehingga mampu menangkap ambiguitas, nuansa emosional, dan ekspresi informal dengan lebih [2][6]. Kemampuan ini sangat krusial dalam menganalisis ulasan pengguna yang bersifat spontan, emosional, dan tidak terstruktur.

Oleh karena itu, penelitian ini dilatarbelakangi oleh kebutuhan mendesak untuk meningkatkan akurasi dan kedalaman analisis sentimen terhadap ulasan aplikasi e-commerce di Indonesia. Penggunaan IndoBERT diharapkan mampu menghasilkan klasifikasi sentimen yang tidak hanya unggul dari sisi akurasi, tetapi juga lebih mencerminkan karakteristik linguistik dan konteks budaya pengguna di Indonesia. Hasilnya akan memberikan wawasan yang lebih bermakna bagi pengembang platform dalam merespons umpan balik pengguna dan meningkatkan kualitas layanan secara berkelanjutan.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui serangkaian langkah inti yang tersusun secara sistematis, dimulai dari import data hingga penarikan kesimpulan. Secara garis besar, seluruh alur proses penelitian yang dilakukan dirangkum secara visual pada diagram alir dalam Gambar 1.



Gambar 1. Alur Penelitian

2.1 Data *Selection* (Pengumpulan Data)

Data yang digunakan dalam penelitian ini bersumber dari data primer yang dikumpulkan langsung melalui teknik web scraping. Proses pengambilan data dilakukan pada tiga platform e-commerce Shopee, Tokopedia, dan Lazada masing-masing sebanyak 4.000 ulasan, sehingga total data yang dihimpun berjumlah 12.000 ulasan. Penerapan metode scraping ini bertujuan untuk memperoleh ulasan pengguna yang asli dan relevan guna mendukung analisis sentimen dalam penelitian [9].

2.2 Labeling

Labeling merupakan proses pemberian kategori pada data ulasan untuk merepresentasikan sentimen yang terkandung di dalam teks. Pada penelitian ini, proses labeling dilakukan secara otomatis berdasarkan skor rating yang diberikan oleh pengguna pada masing-masing ulasan. Skema labeling yang digunakan dibagi menjadi tiga kelas sentimen, yaitu negatif (label 0) untuk ulasan dengan skor 1 dan 2, netral (label 1) untuk ulasan dengan skor 3, serta positif (label 2) untuk ulasan dengan skor 4 dan 5. Pendekatan ini umum digunakan dalam penelitian analisis sentimen e-commerce karena bersifat objektif dan mudah direproduksi [5]. Namun demikian, metode labeling berbasis rating memiliki keterbatasan, khususnya pada kasus di mana teks ulasan tidak sepenuhnya selaras dengan skor rating yang diberikan pengguna. Meskipun demikian, pendekatan ini tetap dipilih sebagai tahap awal pembentukan data latih agar model IndoBERT dapat mempelajari pola bahasa sentimen secara umum pada masing-masing kelas.

2.3 *Preprocessing* Data

Tahap *preprocessing* data merupakan langkah awal yang bertujuan menata serta mempersiapkan teks sebelum dianalisis oleh model, sehingga hasil pemrosesan dapat lebih optimal [10]. Dalam proses pengambilan data, umumnya

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

masih ditemukan teks yang tidak tertata dan berisi beragam karakter yang tidak diperlukan. Pada data yang dikumpulkan, sering ditemukan teks yang tidak teratur serta mengandung berbagai karakter yang tidak dibutuhkan. Oleh sebab itu, preprocessing berfungsi untuk meminimalkan gangguan atau noise pada data [11]. Pada penelitian ini, tahap *preprocessing* mencakup *Data Cleaning, Case Folding,Data Filtering*, serta *Balancing*.

a. Data *Cleaning*
   Data *Cleaning* adalah proses pembersihan dokumen teks dari elemen-elemen yang tidak berhubungan dengan informasi utama yang ingin dianalisis. Pada tahap ini, komponen seperti URL, tanda pagar (hashtag), nama pengguna (@username), alamat email, emotikon (misalnya :@, :*, :D), serta berbagai tanda baca seperti koma, titik, atau simbol lain akan dihapus agar teks lebih rapi dan terstruktur [12]. Tahap *Data Cleaning* penting dilakukan untuk memastikan bahwa data yang diolah tidak mengandung gangguan yang dapat memengaruhi kualitas analisis.
b. *Case Folding*
   Case folding merupakan proses mengubah seluruh karakter dalam teks menjadi huruf kecil guna menjaga konsistensi penulisan [13]. Langkah *Case Folding* dilakukan agar model tidak menganggap huruf kapital dan huruf kecil sebagai entitas berbeda, sehingga dapat mengurangi variasi kata yang tidak diperlukan [14].
c. Data *Filtering*
   Data filtering merupakan proses penyaringan terhadap kata-kata yang tidak memiliki nilai informasi atau tidak relevan, sehingga hanya istilah yang bermakna dan mendukung analisis yang dipertahankan [15]. Tahapan ini penting dilakukan untuk mengurangi noise dalam data serta memastikan bahwa teks yang dianalisis benar-benar mencerminkan konteks sebenarnya.
d. *Balancing*
   Balancing merupakan proses penyeimbangan distribusi kelas dalam sebuah dataset agar jumlah sampel pada tiap kategori tidak timpang atau didominasi oleh salah satu kelas. Langkah ini penting dilakukan untuk mencegah bias pada model, terutama ketika data mayoritas jauh lebih banyak dibandingkan data minoritas [16].

Ketidakseimbangan kelas diketahui sebagai salah satu faktor utama yang memengaruhi kinerja model klasifikasi, karena algoritma cenderung mengoptimalkan akurasi keseluruhan dengan mengabaikan kelas minoritas, sehingga menghasilkan prediksi yang tidak andal pada kelompok tersebut. Salah satu pendekatan yang dapat diterapkan adalah random oversampling, yaitu teknik peningkatan jumlah sampel kelas minoritas melalui replikasi acak dari instans yang sudah ada [17]. Berbeda dengan metode sintetik seperti SMOTE, random oversampling tidak menghasilkan data baru berdasarkan interpolasi fitur, melainkan menyalin ulang data minoritas secara langsung. Meskipun sederhana, teknik ini efektif dalam meningkatkan proporsi kelas minoritas sehingga distribusi kelas menjadi lebih seimbang dan memungkinkan model belajar dengan representasi yang lebih adil terhadap seluruh kategori.

2.4 Tokenization using IndoBERT Tokenizer

Sebelum data digunakan pada tahap pelatihan model, teks terlebih dahulu diproses melalui tokenisasi menggunakan tokenizer IndoBERT. Proses tokenisasi ini mengubah setiap bagian teks menjadi bentuk numerik sehingga dapat diproses dan dipahami oleh model secara efektif [18].

2.5 Fine Tuning IndoBERT Model

Fine-tuning merupakan proses pelatihan ulang model yang sudah pre-trained menggunakan dataset yang lebih kecil namun relevan dengan tugas tertentu, seperti analisis sentimen. Teknik ini memungkinkan model, yang sebelumnya telah memahami bahasa secara umum, untuk menyesuaikan diri dengan kebutuhan spesifik. Meskipun IndoBERT telah dilatih dengan data bahasa Indonesia, model ini masih bersifat umum, sehingga fine-tuning diperlukan agar performanya optimal pada konteks tertentu [19].

Pada penelitian ini, fine-tuning dilakukan menggunakan model IndoBERT untuk tugas klasifikasi sentimen tiga kelas (negatif, netral, dan positif). Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratified sampling untuk menjaga proporsi kelas. Proses fine-tuning dilakukan selama satu epoch dengan ukuran batch sebesar 8, weight decay sebesar 0.01, serta warm up sebanyak 10 langkah. Evaluasi model dilakukan pada setiap akhir epoch menggunakan metrik accuracy dan macro F1-score untuk mengakomodasi potensi ketidakseimbangan kelas. Model terbaik dipilih secara otomatis berdasarkan performa evaluasi dengan mekanisme load best model at end. Konfigurasi ini dipilih dengan mempertimbangkan keterbatasan ukuran data dan efisiensi komputasi, sekaligus untuk mencegah overfitting pada dataset berukuran kecil.

2.6 Evaluation

Setelah proses *fine-tuning* selesai, model dievaluasi menggunakan *confusion matrix*, yaitu alat yang digunakan untuk mengukur kinerja model klasifikasi dengan memperlihatkan perbandingan antara hasil prediksi dan label sebenarnya dalam format matriks. Matriks ini terdiri dari beberapa komponen, yaitu *True Positive (TP), True Negative (TN), False Positive (FP),* dan *False Negative (FN).* Nilai-nilai tersebut kemudian dimanfaatkan untuk menghitung berbagai metrik evaluasi seperti akurasi, recall, presisi, dan F1-score [20].

a. *Accuracy*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

Dalam analisis sentimen terhadap ulasan e-commerce seperti Shopee, Tokopedia, dan Lazada, akurasi (accuracy) menggambarkan seberapa tepat model dalam mengklasifikasikan ulasan ke dalam sentimen positif, negatif, maupun netral. Nilai akurasi diperoleh berdasarkan perhitungan yang ditunjukkan pada Formula 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

b. *Precission*
Mengacu pada Formula 2, presisi (precision) digunakan untuk mengukur seberapa tepat prediksi model untuk setiap kelas sentimen. Dalam kasus multi-kelas, presisi dihitung untuk masing-masing kelas (positif, negatif, netral) dan kemudian dapat digabungkan menggunakan macro-averaging atau weighted-averaging.

$$Precission = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall*
Recall menunjukkan kemampuan model dalam mengenali seluruh ulasan yang benar-benar termasuk pada setiap kelas sentimen. Pada Formula 3, *recall* mengukur seberapa lengkap model dalam mendeteksi semua instance dari kelas tertentu.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. *F1-score*
F1-Score merupakan matrik evaluasi yang menggabungkan presisi dan recall menjadi satu nilai tunggal, sehingga memberikan keseimbangan antara kedua metrik tersebut. F1-Score dihitung sebagai rata-rata harmonik dari presisi dan recall, sehingga memastikan bahwa baik akurasi prediksi maupun kemampuan mendeteksi semua instance kelas sama-sama diperhitungkan dalam evaluasi akhir. Perumusannya disajikan pada Formula 4

$$F1 - Score = \frac{Precission\ x\ Recall}{Precission+Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Penelitian ini memaparkan hasil penerapan model klasifikasi sentimen berbasis IndoBERT pada ulasan pengguna aplikasi e-commerce (Shopee, Tokopedia, dan Lazada) yang diperoleh dari Google Play Store. Penyajian hasil disusun secara sistematis sehingga setiap langkah penelitian, mulai dari tahap preprocessing hingga evaluasi performa model, dapat dipahami dengan jelas dan ditelusuri secara runtut.

3.1 Pengumpulan Data

Penelitian ini memanfaatkan data berupa ulasan pengguna dari aplikasi e-commerce Shopee, Tokopedia, dan Lazada yang diperoleh melalui Google Play Store. Proses pengambilan data (scraping) menghasilkan 1.500 ulasan dalam bentuk dataset mentah. Dataset tersebut memuat tujuh kolom (fitur), dengan karakteristik lengkapnya ditampilkan pada Tabel 1.

Tabel 1. Dataset Penelitian

App	Username	Score	Content	Review Created	Version
Shopee	Ros Salinda	4	bagus dan baik	…	3.60.30
Shopee	Friyoan Suryo	5	Sangat membantu, jika kita punya keperluan tinggal dibeli saja di aplikasi ini. Pokoknya sangat membantu.	…	3.59.41
Shopee	ambon rege	5	mantap	…	…
Shopee	Dedi Suhendra	5	good	…	…
Shopee	Sandi Uno	5	bagus	…	3.59.41
Shopee	…	…	…	…	3.57.31
Shopee	…	…	…	…	…
Shopee	…	…	…	…	…
Shopee	…	…	…	…	…
Lazada	Mochamad Nurdin	5	…	…	7.85.0
Lazada	widiastuti wisiastuti	5	ok	…	7.85.0
Lazada	Ircham Maulana	4	mantap	…	7.86.0
Lazada	Estu Hayatul	4	suka belanja di lazada	…	7.85.0

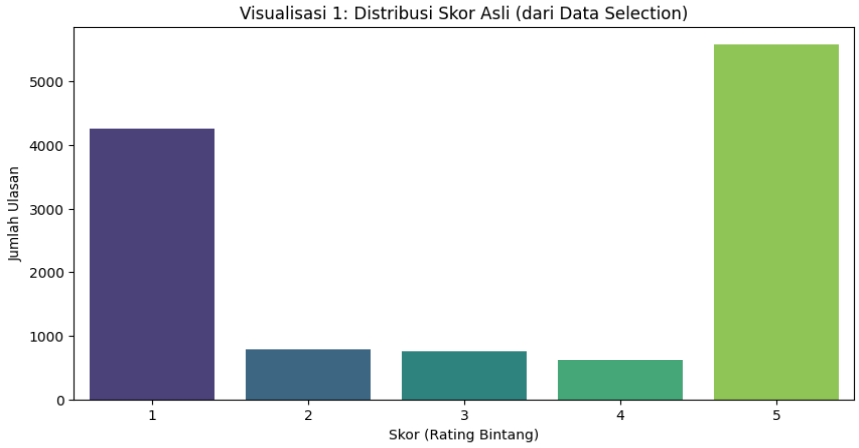
Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

App	Username	Score	Content	Review Created	Version
Lazada	Saiful taurus	5	Aplikasi Lazada bagus banget, is the best pokoknya	…	7.85.0

Dari 7 fitur tersebut, penelitian ini berfokus pada dua fitur utama untuk analisis sentimen, yaitu content sebagai data teks yang akan dianalisis dan score sebagai dasar untuk pelabelan sentimen.

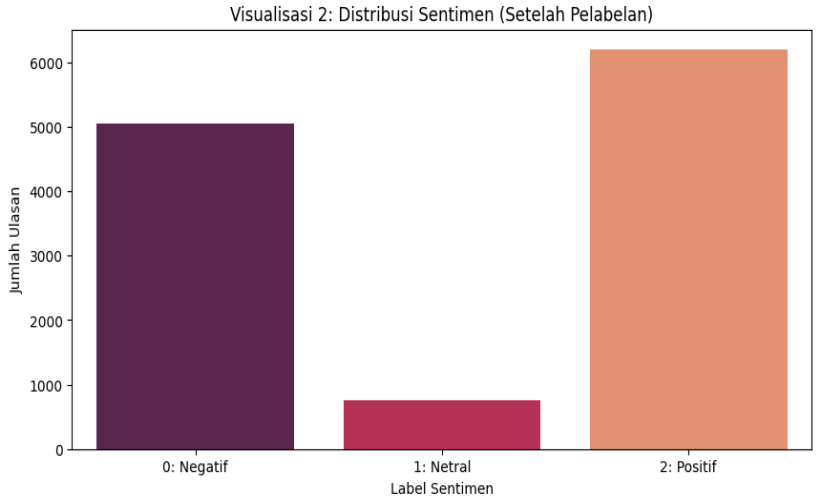
3.2 Labeling

Labeling merupakan proses pemberian kategori pada data ulasan untuk merepresentasikan sentimen yang terkandung di dalam teks. Pada penelitian ini, proses labeling dilakukan secara otomatis berdasarkan skor rating yang diberikan oleh pengguna pada masing-masing ulasan. Skema labeling yang digunakan dibagi menjadi tiga kelas sentimen, yaitu negatif (label 0) untuk ulasan dengan skor 1 dan 2, netral (label 1) untuk ulasan dengan skor 3, serta positif (label 2) untuk ulasan dengan skor 4 dan 5. Pendekatan ini umum digunakan dalam penelitian analisis sentimen e-commerce karena bersifat objektif dan mudah direproduksi [5]. Namun demikian, metode labeling berbasis rating memiliki keterbatasan, khususnya pada kasus di mana teks ulasan tidak sepenuhnya selaras dengan skor rating yang diberikan pengguna. Meskipun demikian, pendekatan ini tetap dipilih sebagai tahap awal pembentukan data latih agar model IndoBERT dapat mempelajari pola bahasa sentimen secara umum pada masing-masing kelas.



Gambar 2. Distribusi Rating

Gambar 2 memperlihatkan bahwa mayoritas ulasan berada pada skor 5 (positif) dan skor 1 (negatif). Setelah proses pelabelan dilakukan, distribusi sentimen yang ditampilkan pada Gambar 3 kembali menegaskan adanya ketidakseimbangan data (imbalance) yang cukup besar.



Gambar 3. Distribusi Rating Setelah Labeling

Data awal terdiri dari 6.201 ulasan positif (Label 2), 5.044 ulasan negatif (Label 0), dan hanya 755 ulasan netral (Label 1). Ketidakseimbangan ini, terutama pada kelas "Netral", dapat menyebabkan model menjadi bias dan tidak mampu memprediksi kelas minoritas secara akurat. Oleh karena itu, diperlukan langkah preprocessing dan penanganan data tidak seimbang.

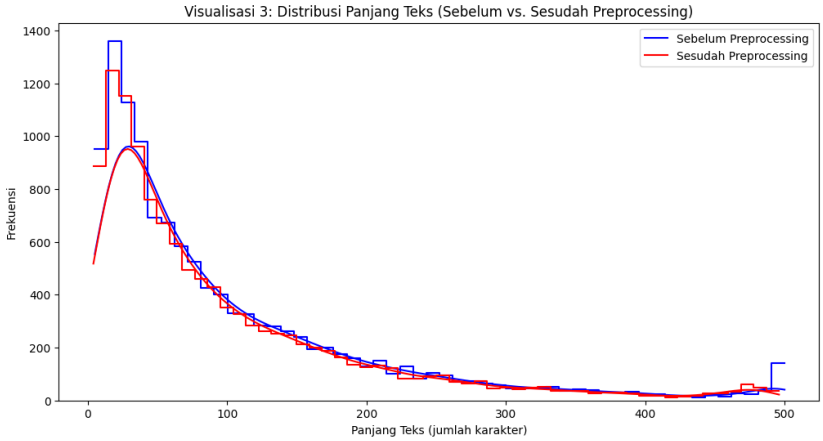
Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

3.3 Data Preprocessing

Tahap *data preprocessing* bertujuan untuk membersihkan teks ulasan dan mengatasi ketidakseimbangan data. Proses ini dibagi menjadi dua sub-tahap.

3.3.1 Data Cleaning, Case Folding, dan Filtering

Teks ulasan terlebih dahulu melalui serangkaian tahap preprocessing yang meliputi data cleaning untuk menghilangkan URL, mention, tagar, tanda baca, serta angka sehingga hanya menyisakan karakter alfabet. Selanjutnya, dilakukan case folding dengan mengubah seluruh teks menjadi huruf kecil. Tahap filtering diterapkan untuk mengidentifikasi entri yang tidak memiliki konten teks setelah proses pembersihan. Namun, hasil preprocessing menunjukkan bahwa tidak terdapat ulasan yang menjadi kosong, sehingga seluruh 12.000 ulasan tetap digunakan dalam tahap analisis selanjutnya. Perbandingan panjang karakter sebelum dan sesudah preprocessing ditampilkan pada Gambar 4.

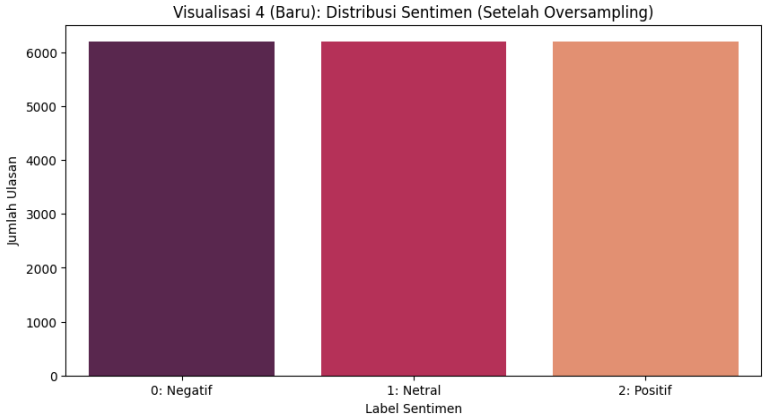


Gambar 4. Distribusi Panjang Teks

3.3.2 *Balancing*

Berdasarkan temuan pada bagian 3.2, distribusi kelas pada data menunjukkan ketidakseimbangan yang cukup signifikan. Untuk mengatasi hal tersebut, penelitian ini menerapkan Random Oversampling, yaitu teknik yang menduplikasi data pada kelas minoritas Negatif dan Netral secara acak hingga jumlahnya setara dengan kelas mayoritas. Sebelum dilakukan oversampling, jumlah data pada masing-masing kelas adalah Negatif 5.044, Netral 755, dan Positif 6.201, dengan target penyamaan pada angka 6.201.

Setelah proses oversampling, ketiga kelas tersebut menjadi seimbang dengan masing-masing berjumlah 6.201 data. Setelah proses oversampling, total dataset yang digunakan untuk pelatihan model bertambah menjadi 18.603 data ulasan yang seimbang. Distribusi data yang telah seimbang ditunjukkan pada Gambar 5. Langkah ini krusial untuk memastikan model mendapatkan porsi data yang adil bagi setiap kelas.



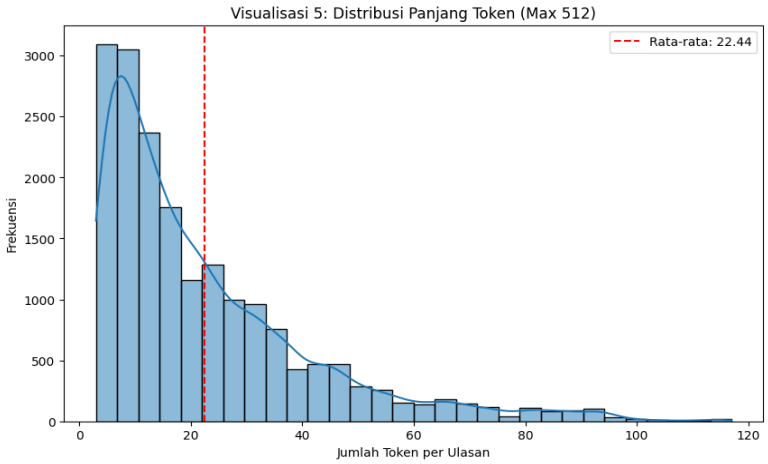
Gambar 5. Distribusi Sentimen Setelah Oversampling

3.4 Tokenisasi IndoBERT

Data yang telah melalui proses pembersihan serta penyeimbangan (sebanyak 18.603 ulasan) selanjutnya ditokenisasi menggunakan tokenizer dari model indobenchmark/indobert-base-p1. Tokenisasi merupakan tahap konversi teks mentah menjadi rangkaian token baik berupa kata maupun sub-kata yang dapat diolah oleh model. IndoBERT

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

mengimplementasikan WordPiece tokenization, suatu mekanisme yang memungkinkan pemecahan kata tidak dikenal (out-of-vocabulary) menjadi unit sub-kata sehingga tetap dapat direpresentasikan dengan baik oleh model. Distribusi panjang token pada seluruh data dapat dilihat pada Gambar 6.



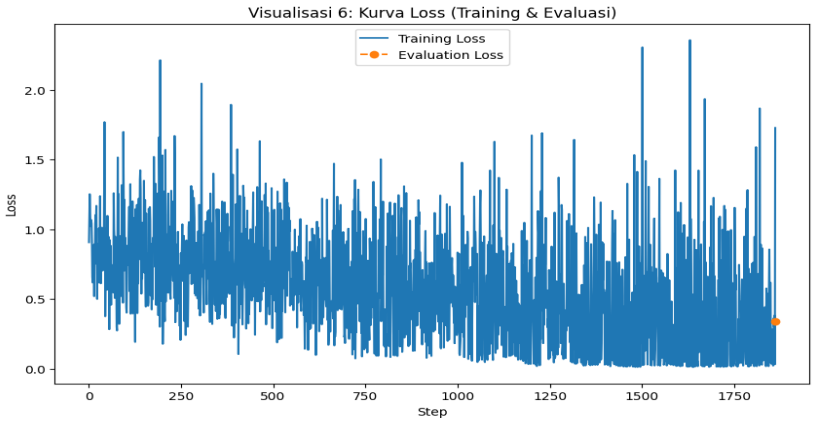
Gambar 6. Distribusi Panjang Token

Distribusi panjang token menunjukkan bahwa rata-rata ulasan memiliki panjang sekitar 22,44 token. Hampir seluruh ulasan berada di bawah 512 token. Berdasarkan temuan ini, panjang sekuens maksimum (max_length) untuk model ditetapkan sebesar 512.

3.5 Modelling

Dataset yang telah melalui proses tokenisasi kemudian dibagi menjadi data pelatihan dan data pengujian dengan rasio 80% dan 20% menggunakan stratified sampling untuk menjaga proporsi kelas sentimen. Tahap pemodelan dilakukan melalui proses fine-tuning model IndoBERT yang telah melalui pre-training sebelumnya, sehingga model dapat disesuaikan untuk tugas klasifikasi sentimen tiga kelas (negatif, netral, dan positif).

Proses fine-tuning dilakukan selama satu epoch dengan ukuran batch sebesar 8 untuk data pelatihan dan data evaluasi. Untuk mengoptimalkan proses pembelajaran, digunakan weight decay sebesar 0.01 serta warmup sebanyak 10 langkah. Evaluasi model dilakukan pada setiap akhir epoch menggunakan metrik accuracy dan macro F1-score guna memberikan gambaran performa yang lebih representatif pada data yang berpotensi tidak seimbang. Model terbaik dipilih secara otomatis berdasarkan hasil evaluasi pada data pengujian dengan mekanisme *load best model at end*.



Gambar 7. Kurva Loss

Pada Gambar 7, kurva Training Loss (biru) menunjukkan tren penurunan yang stabil, hal ini menunjukkan bahwa model mampu memahami pola yang terdapat pada data latih. Kurva Evaluation Loss (garis oranye) memperlihatkan nilai loss yang sangat kecil pada akhir proses evaluasi, yaitu 0.3376. Nilai ini menunjukkan bahwa model tidak mengalami overfitting dan memiliki kemampuan generalisasi yang baik.

3.6 Evaluasi

Evaluasi akhir dilakukan menggunakan 3.721 data uji untuk menilai performa model setelah proses fine-tuning. Mengingat dataset telah diseimbangkan melalui teknik oversampling sehingga jumlah data pada setiap kelas relatif setara, metrik evaluasi utama yang digunakan dalam penelitian ini adalah Macro F1-Score, karena memberikan bobot

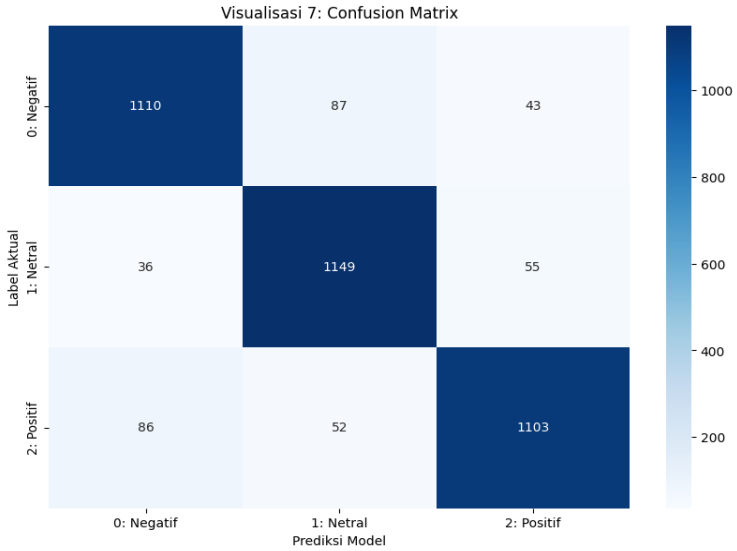
Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

yang sama pada setiap kelas. Berdasarkan hasil evaluasi, model IndoBERT memperoleh nilai akurasi sebesar 90% dengan Macro Precision, Macro Recall, dan Macro F1-Score masing-masing sebesar 0.90, sebagaimana ditampilkan pada Tabel 2.

Tabel 2. Matrik Evaluasi

Kelas	Precision	Recall	F1-Score
0 = Negatif	0.90	0.90	0.90
1 = Netral	0.89	0.93	0.91
2 = Positif	0.92	0.89	0.90
Accuracy	0.90	0.90	0.90
Rata-rata	0.90	0.90	0.90

Berdasarkan Tabel 2, kinerja model menunjukkan hasil yang sangat baik dan seimbang di seluruh kelas sentimen, yang tercermin dari nilai Macro F1-Score sebesar 0.90. Penggunaan Macro Average menegaskan bahwa performa model konsisten pada setiap kelas tanpa dipengaruhi oleh dominasi kelas tertentu. Kelas Netral menunjukkan performa paling stabil dengan nilai Precision sebesar 0.89 dan Recall sebesar 0.93, menghasilkan F1-Score sebesar 0.91. Sementara itu, kelas Negatif dan Positif masing-masing memperoleh F1-Score sebesar 0.90, menunjukkan bahwa model mampu membedakan polaritas sentimen secara konsisten pada ketiga kelas.



Gambar 8. Confusion Matrix

Confusion Matrix pada Gambar 8 memberikan gambaran rinci mengenai performa klasifikasi model IndoBERT pada 3.721 data uji. Model mampu mengklasifikasikan dengan benar sebanyak 1.110 ulasan Negatif, 1.149 ulasan Netral, dan 1.103 ulasan Positif, yang ditunjukkan oleh nilai pada diagonal utama matriks. Kesalahan klasifikasi relatif kecil dan tersebar antar kelas. Pada kelas Negatif, sebanyak 87 ulasan salah diprediksi sebagai Netral dan 43 ulasan sebagai Positif. Untuk kelas Netral, terdapat 36 ulasan yang keliru diklasifikasikan sebagai Negatif dan 55 ulasan sebagai Positif. Sementara itu, pada kelas Positif, sebanyak 86 ulasan diprediksi sebagai Negatif dan 52 ulasan sebagai Netral. Secara umum, dominasi nilai pada diagonal matriks menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dan seimbang pada ketiga kelas sentimen. Pola kesalahan yang relatif merata ini mengindikasikan bahwa model tidak bias terhadap kelas tertentu, sejalan dengan nilai Macro F1-Score yang tinggi dan mengkonfirmasi kemampuan generalisasi model pada data uji.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengatasi keterbatasan pendekatan tradisional dalam analisis sentimen ulasan e-commerce berbahasa Indonesia, khususnya terkait ketidakseimbangan kelas dan ketidakmampuan representasi berbasis TF-IDF dalam menangkap konteks linguistik. Hasil penelitian menunjukkan bahwa penerapan model IndoBERT yang dikombinasikan dengan teknik Random Oversampling mampu menghasilkan performa klasifikasi yang seimbang pada ketiga kelas sentimen, tercermin dari nilai Macro F1-Score sebesar 0.90 pada 3.721 data uji. Penggunaan Macro F1-Score menegaskan bahwa peningkatan performa tidak didominasi oleh kelas mayoritas, melainkan juga mencerminkan kemampuan model dalam mengenali kelas minoritas, terutama kelas Netral yang pada penelitian sebelumnya sering mengalami penurunan performa. Temuan ini mengonfirmasi bahwa pendekatan berbasis Transformer lebih efektif dalam menangani kompleksitas bahasa alami pengguna Indonesia dan mengatasi keterbatasan metode machine learning tradisional pada skenario data yang tidak seimbang. Meskipun demikian,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2127−2135
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8745

penelitian ini masih memiliki keterbatasan pada ukuran dan keberagaman dataset. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi teknik penyeimbangan data alternatif, membandingkan performa dengan model Transformer lain seperti IndoBERT-lite atau IndoGPT, serta memperluas cakupan data agar model mampu menangani variasi bahasa yang lebih kompleks dan dinamis.

REFERENCES

[1] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," J. Inform. J. Pengemb. IT, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.

[2] I. S. Milal, M. Hasanudin, M. A. N. Azhari, R. A. Nugraha, N. Agustina, and S. E. Damayanti, "Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma SVM," NARATIF J. Ilm. Nas. Ris. Apl. dan Tek. Inform., vol. 05, no. 01, pp. 34–45, 2023. [Online]. Available: https://naratif.utb-univ.ac.id/index.php/naratif/article/download/191/96.

[3] T. Puspa, R. Sanjaya, A. Fauzi, A. Fitri, and N. Masruriyah, "Analisis Sentimen Ulasan Pada E-Commerce Shopee Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," INFOTECH J. Inform. Teknol., vol. 4, no. 2722–9386, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.

[4] J. Loso, H. Susila, U. Najirah, and I. K. S. Satwika, "Transformasi Digital (Teori & Implementasi Menuju Era Society 5.0)", E. Rianty, Ed.1 Bekasi, Indonesia: Son Pedia Publishing Indonesia, May 2024. [Online]. Available:https://www.researchgate.net/publication/380462238_TRANSFORMASI_DIGITAL_Teori_implementasi_Menuju_Era_Society_50

[5] S. Adryan, N. Firman, and R. Aviv Yuniar, "Analisis sentimen aplikasi shopee, tokopedia, lazada dan blibli menggunakan leksikon dan random forest," JITET (Jurnal Inform. dan Tek. Elektro Ter., vol. 12, no. 3, pp. 3576–3587, 2024, doi: 10.23960/jitet.v12i3S1.5155.

[6] Steven and F. Indah, "Analisis Sentimen Membandingkan Pengguna Aplikasi E-Commerce Tokopedia Dan Shopee Menggunakan Algoritma Naive Bayes," pp. 32–39, 2024. [Online]. Available: https://jurnal.ubd.ac.id/index.php/poters/article/view/3588.

[7] A. G. F. N. Pramita, "Aplikasi Lazada Menggunakan Metode Naive Bayes," J. Digit, vol. 14, no. 1, pp. 23–30, 2024, doi: 10.51920/jd.v14i1.362.

[8] B. Z. Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," J. Appl. Informatics Comput., vol. 6, no. 2, pp. 220–225, 2022, doi: 10.30871/jaic.v6i2.4725.

[9] G. T. Fadilah, L. Muflikhah, and R. S. Perdana, "Analisis Sentimen Produk Hijab Pada E-Commerce Tokopedia Menggunakan Algoritma Support Vector," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 9, no. 2, pp. 1–9, 2025. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14390.

[10] A. Fayola and R. Darianty, "Analisis Sentimen Ulasan Produk Kecantikan di Tokopedia Menggunakan IndoBERT," ISAJ (Information Syst. Appl. Journal), vol. 01, no. 01, pp. 11–18, 2025, doi: https://doi.org/isaj.v1i1.1.

[11] R. S. Nuraini, "Analisis Sentimen Pengguna Aplikasi Agoda Di Google Play Store Menggunakan Algoritma Naiva Bayes," J. Inf. Syst. Manag., vol. 7, no. 1, pp. 24–29, 2025, doi: 10.24076/joism.2025v7i1.2066.

[12] A. J. Putri, A. S. Syafira, and M. E. Purbaya, "Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine," J. TRINISTIK, vol. 01, no. 1, pp. 16–21, 2022, doi: 10.20895/trinistik.v1i1.447.

[13] M. Xanderina et al., "Analisis Sentimen Ulasan E-Commerce Shopee Pada Google Play Store Menggunakan Machine Learning," J-ENSISTEC (Journal Eng. Sustain. Technol., vol. 10, no. 02, pp. 990–998, 2024, doi: 10.31949/jensitec.v10i02.9071.

[14] E. A. Junita and R. R. Suryono, "Analisis sentimen hate speech mengenai calon wakil presiden indonesia menggunakan algoritma bert," JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 9, no. 4, pp. 2042–2053, 2024, doi: 10.29100/jipi.v9i4.5544.

[15] A. V. Utomo, A. Setiawan, and M. Arifin, "Analisis Sentimen Ulasan Hijab Aulia dengan Metode Support Vector Machine untuk Kepuasan Pelanggan," JUSIBI (JURNAL Sist. Inf. DAN E-BISNIS), vol. 7, no. 2655–7541, pp. 85–95, 2025, doi:10.54650/jusibi.v7i2.607.

[16] I. Irma Surya Kumala, M. Yasin Aril, and S. Irvan Abraham, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," Jambura J. Electr. Electron. Eng., vol. 5, pp. 32–35, 2023, doi:10.37905/jjeee.v5i1.16830.

[17] R. F. Rahmanda, Y. Sibaroni, and S. S. Prasetiyowati, "Effectiveness of Bi-GRU and FastText in Sentiment Analysis of Shopee App Reviews," Rayhan Fadhil, Rahmanda Yuliant, Sibaroni Sri Suryani, Prasetiyowati, vol. 9, no. 1, pp. 444–454, 2025, doi:10.33395/sinkron.v9i1.14474.

[18] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning , Logistic Regression , and Linear SVM," J. Appl. Informatics Comput., vol. 9, no. 5, pp. 2616–2628, 2025, doi: 10.30871/jaic.v9i5.10537.

[19] B. Eka, and S. Dewi, "Model Analisis Sentimen Pada Kendaraan Listrik Menggunakan Algoritma Indobertweet Dan Indobert," ANTIVIRUS J. Ilm. Tek. Inform., vol. 19, no. 1, pp. 169–179, 2025, doi: 10.35457/antivirus.v19i1.4416 169.

[20] M. G. Al-kadzim, "Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT," Digit. Transform. Technol., vol. 4, no. 2, pp. 1167–1174, 2024, doi: 10.47709/digitech.v4i2.5312.