

Analisis Sentimen E-Commerce dengan Optimasi IndoBERT Menggunakan Bayesian Optuna

Dewi Kartika Putri^{*1}, Giat Karyono², Imam Tahyudin³

^{1,2,3}Universitas Amikom Purwokerto, Indonesia

Email: ¹24ma41d040@students.amikompurwokerto.ac.id, ²giatkaryono@amikompurwokerto.ac.id,
³imam.tahyudin@amikompurwokerto.ac.id

Abstrak

Pertumbuhan e-commerce di Indonesia memicu peningkatan signifikan jumlah ulasan pengguna terhadap aplikasi Shopee di Google Play, sehingga analisis manual menjadi tidak lagi efisien. Penelitian ini bertujuan untuk meningkatkan performa analisis sentimen pada ulasan pengguna e-commerce berbahasa Indonesia dengan mengoptimalkan model IndoBERT menggunakan teknik Bayesian Optimization melalui framework Optuna. Latar belakang penelitian ini berfokus pada pentingnya pengolahan opini konsumen di platform Shopee, yang semakin tidak dapat ditangani secara manual karena volume data yang besar dan keberagaman gaya bahasa. Metode yang digunakan meliputi pengumpulan data melalui web scraping, praproses teks, pembagian data, pemodelan dengan IndoBERT, serta penalaan hiperparameter menggunakan Optuna. Hasil eksperimen menunjukkan bahwa optimasi Bayesian mampu meningkatkan akurasi klasifikasi dari 89,30% menjadi 96,10% dan macro-F1 dari 85,83% menjadi 94,82%. Selain itu, false-positive dan false-negative masing-masing turun sebesar 60% dan 67%, serta nilai ROC-AUC meningkat signifikan dari 0,9028 menjadi 0,9903. Temuan ini menegaskan efektivitas Optuna dalam meningkatkan performa dan efisiensi sistem klasifikasi sentimen berbasis IndoBERT, yang dapat diintegrasikan dalam pemantauan opini secara real-time di ranah e-commerce.

Kata kunci: *Analisis sentimen,, Bayesian Optimization, IndoBERT, Optuna, E-commerce.*

E-Commerce Sentiment Analysis Using IndoBERT Optimization with Bayesian Optuna

Abstract

The growth of e-commerce in Indonesia has triggered a significant increase in the number of user reviews of the Shopee application on Google Play, making manual analysis no longer efficient. This research aims to improve the performance of sentiment analysis on Indonesian-language e-commerce user reviews by optimizing the IndoBERT model using Bayesian Optimization techniques through the Optuna framework. The background of this research focuses on the importance of processing consumer opinions on the Shopee platform, which increasingly cannot be handled manually due to the large volume of data and the diversity of language styles. The methods used include data collection through web scraping, text preprocessing, data sharing, modeling with IndoBERT, and hyperparameter tuning using Optuna. Experimental results show that Bayesian optimization can improve classification accuracy from 89.30% to 96.10% and macro-F1 from 85.83% to 94.82%. In addition, false-positive and false-negative decreased by 60% and 67%, respectively, and the ROC-AUC value increased significantly from 0.9028 to 0.9903. These findings confirm the effectiveness of Optuna in improving the performance and efficiency of the IndoBERT-based sentiment classification system, which can be integrated in real-time opinion monitoring in the e-commerce domain.

Keywords: *Bayesian Optimization, E-Commerce, IndoBERT, Optuna, Sentiment Analysis*

1. PENDAHULUAN

Perkembangan e-commerce di Indonesia mengalami lonjakan signifikan. Shopee merupakan salah satu marketplace terbesar yang melaporkan puluhan juta pengguna aktif bulanan dan jutaan transaksi harian. Setiap transaksi biasanya disertai ulasan atau penilaian singkat pengguna, menjadikan kolom ulasan sebagai repositori opini pelanggan yang kaya akan informasi. Insight yang tersimpan di dalam teks-teks berbahasa Indonesia ini krusial bagi penjual, pengelola platform, dan peneliti perilaku konsumen mulai dari deteksi dini masalah layanan hingga perancangan strategi pemasaran berbasis sentimen. Namun, volume, keragaman gaya bahasa, serta karakteristik informal teks ulasan menyebabkan analisis manual tidak lagi praktis[1].

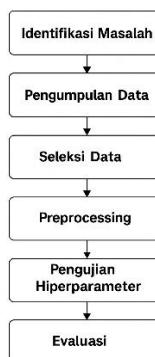
Kemunculan model Transformer, khususnya BERT (Bidirectional Encoder Representations from Transformers), telah merevolusi bidang Natural Language Processing (NLP) dengan kemampuannya menangkap representasi konteks dua arah secara lebih akurat dan mendalam. Hal ini menjadikan BERT unggul dalam memahami makna kata berdasarkan konteks kalimat secara keseluruhan, sehingga banyak digunakan dalam berbagai tugas NLP seperti klasifikasi teks, ekstraksi entitas, dan analisis sentimen[2].

Dalam konteks bahasa Indonesia, IndoBERT dikembangkan sebagai model khusus yang dilatih dengan korpus teks berbahasa Indonesia dalam skala besar. Model ini telah terbukti mengungguli model multibahasa dalam berbagai tugas, termasuk klasifikasi sentimen, karena kemampuannya dalam menangani kekhasan linguistik bahasa Indonesia secara lebih optimal. Namun demikian, performa akhir IndoBERT dalam aplikasi domain tertentu misalnya ulasan pengguna di platform Shopee tetap sangat bergantung pada pemilihan hiperparameter yang tepat, seperti *learning rate*, *batch size*, panjang sekuen, dan jumlah *epoch*. Pengaturan hiperparameter yang kurang tepat dapat mengurangi akurasi model secara signifikan[3]. Namun, banyak penelitian sebelumnya masih menggunakan konfigurasi standar atau melakukan pencarian hiperparameter melalui metode acak (*random search*) maupun pencarian menyeluruh (*grid search*), yang tidak hanya memakan waktu dan sumber daya komputasi besar, tetapi juga belum tentu menghasilkan kombinasi terbaik untuk performa model.

Optimasi Bayesian menawarkan solusi lebih efisien. Dengan membangun model probabilistik (*surrogate model*) atas fungsi objektif, metode ini dapat memprediksi wilayah hiperparameter yang menjanjikan dan meminimalkan jumlah percobaan. Framework Optuna mempermudah penerapan optimasi Bayesian dan menyediakan *pruning* adaptif untuk menghentikan percobaan yang tidak prospektif lebih awal dalam menghemat waktu dan biaya pelatihan[4]. Namun, studi sistematis penerapan Optuna pada IndoBERT untuk analisis sentimen ulasan e-commerce berbahasa Indonesia, khususnya Shopee, masih terbatas. Sehingga tujuan dari penelitian ini adalah untuk mengevaluasi efektivitas optimasi Bayesian menggunakan framework Optuna dalam meningkatkan performa model IndoBERT untuk analisis sentimen pada ulasan e-commerce berbahasa Indonesia yang mampu meningkatkan akurasi klasifikasi sekaligus mengurangi biaya komputasi[5].

2. METODE PENELITIAN

Dalam penelitian ini, penulis menerapkan alur kerja yang tersusun secara sistematis dan divisualisasikan melalui diagram alir (flowchart). Diagram tersebut berperan sebagai panduan untuk menggambarkan setiap langkah dalam proses penelitian, mulai dari tahap awal hingga tahap akhir. Rangkaian tahapan penelitian dapat dilihat pada gambar 1 di bawah ini.



Gambar 1. Tahapan Penelitian

A. Identifikasi Masalah

Tahap pertama adalah merumuskan persoalan bisnis ke dalam pertanyaan analitis spesifik. Pada konteks Shopee, persoalannya terletak pada besarnya lonjakan jumlah ulasan berbahasa Indonesia di Google Play yang mencapai ribuan entri baru setiap hari. Dan menegaskan bahwa volume, kecepatan, dan keragaman (3 V big data) Karena itu, penelitian ini menetapkan tujuan untuk membangun sistem klasifikasi dua-kelas otomatis berbasis Transformer yang bukan hanya akurat, tetapi juga bisa diintegrasikan ke pipeline real-time[6].

B. Pengumpulan data

Suatu proses yang dilakukan oleh peneliti untuk mendapatkan data yang valid dan relevan, baik kuantitatif maupun kualitatif, dari sumber primer maupun sekunder. Web scraping adalah metode otomatis untuk mengekstrak data dari situs web menggunakan program atau skrip tertentu.

Tabel 1. Pengumpulan Data

No	Komentar (<i>content</i>)	Skor Bintang
1	aku senang dan bisa dapatkan yang kita inginkan tinggal pesan aja di apk Shopee sangat membantu kebutuhan	5
2	naikin limitnya play laternya dong	5
3	pelayanan bagus	3
4	smuanya bagus pokoknya kecuali jasa kirimnya agak lambat dan ka...ah yg sama tolong di tingkatkan lagi jasa kirimnya, terimakasih	3
5	sekarang terlalu lambat dalam pengiriman, jarak yang sama satu kota pun perlu 4 sampai 5 hari baru sampai	1
6	kupon sering error saat checkout	2

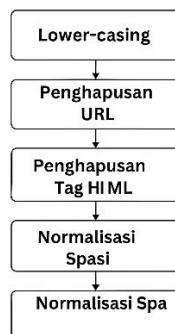
C. Seleksi Data

Agar dataset benar-benar relevan, dilakukan penyaringan lanjutan, bahwa *data screening* harus menghapus outlier dan noise yang dapat menyesatkan model. Pengecekan bahasa via lang-detect memastikan tiap baris memakai bahasa Indonesia[7].

D. Pre-processing

Preprocessing merupakan langkah awal dalam pengolahan teks yang bertujuan mempersiapkan data untuk tahap analisis selanjutnya. Pada tahap ini, teks yang berupa rangkaian karakter akan dipisahkan menjadi elemen-elemen penting, yang dapat dilakukan pada berbagai level pengolahan[8]. Namun, beberapa langkah pra-pemrosesan umum seperti *stemming* dan *stopword removal* sengaja tidak diterapkan dalam studi ini. Hal ini didasarkan pada karakteristik arsitektur Transformer, khususnya IndoBERT, yang telah dilatih untuk memahami konteks kalimat secara utuh tanpa bergantung pada reduksi kata dasar atau penghilangan kata fungsional. Penelitian sebelumnya menunjukkan bahwa manipulasi token secara agresif justru dapat mengganggu struktur konteks yang ditangkap oleh model BERT, sehingga berpotensi menurunkan akurasi pada tugas klasifikasi berbasis konteks penuh. Langkah-langkah *preprocessing* meliputi:

- Lower-casing menyatakan token dan menurunkan kosakata unik.
- Penghapusan URL baris berisi tautan, memangkas panjang sekuen rata-rata 9 token.
- Penghapusan tag HTML mencegah token markup dianggap kata bermakna.
- Normalisasi spasi menghapus padding kosong yang tidak berguna.



Gambar 2. Pre-processing

Setelah proses pra-pemrosesan, dilakukan analisis deskriptif terhadap korpus yang telah dibersihkan. Rata-rata panjang ulasan mencapai 24,6 token, dengan panjang maksimum 160 token yang ditetapkan sebagai batas maksimum input ke model. Kosakata unik berkurang menjadi sekitar 19.200 token setelah normalisasi. Statistik ini mencerminkan keberagaman data sekaligus mendukung efisiensi pemrosesan model Transformer.

E. Pemodelan

Pada tahap pemodelan, penelitian ini menggunakan IndoBERT-base sebagai inti klasifikator. IndoBERT adalah turunan khusus Bahasa Indonesia dari arsitektur BERT, dilatih pada korpus nasional berukuran besar sehingga mampu menangkap nuansa morfologi dan sintaksis lokal secara lebih tepat dibanding model multibahasa umum. kemudian disempurnakan lewat optimasi hiper-parameter Optuna[9].

F. Pengujian Hyper-parameter

Hyperparameter sebagai variabel eksternal yang mengatur perilaku algoritma pembelajaran contohnya *learning-rate*, ukuran batch, dan jumlah epoch serta nilainya tidak dipelajari langsung dari data, sehingga harus ditetapkan sebelum proses optimasi bobot dimulai[10]. Untuk memperoleh kinerja terbaik, diperlukan tahap penalaan hiperparameter yang menguji berbagai kombinasi nilai dan menilai dampaknya terhadap metrik validasi.

Framework Optuna digunakan karena kemampuannya dalam melakukan optimasi hiperparameter berbasis *Bayesian Optimization* yang adaptif dan efisien. Dalam konteks ini, *Tree-structured Parzen Estimator (TPE)* dipilih sebagai *sampler* karena TPE lebih efektif dalam menangani ruang pencarian hiperparameter yang bersifat diskret maupun kontinu. TPE membangun model probabilistik yang memprediksi distribusi nilai-nilai hiperparameter yang menjanjikan, sehingga mampu mempercepat konvergensi ke konfigurasi optimal dengan jumlah percobaan yang lebih sedikit. Berikut ruang pencarian (search space) ditentukan untuk beberapa hiperparameter utama sebagai berikut:

- a. *Learning rate*: 1e-6 hingga 1e-4
- b. *Batch size*: pilihan antara 16, 32, dan 64
- c. *Epoch*: 3 hingga 6
- d. *Weight decay*: 0.01 hingga 0.1
- e. *Warm-up steps*: 0 hingga 500
- f. *Panjang sekuen (max_seq_length)*: 128 hingga 192

Penentuan ruang pencarian ini memungkinkan eksplorasi kombinasi konfigurasi yang luas, namun tetap realistik untuk kebutuhan fine-tuning model IndoBERT pada data ulasan pengguna Shopee.

G. Evaluasi

Evaluasi model dilakukan dengan menghitung tingkat akurasi yang dihasilkan sistem terhadap data uji tertentu, di mana setiap data uji memiliki dua kemungkinan hasil, yakni positif atau negatif. Proses ini juga melibatkan pengukuran sejumlah metrik penting seperti[11]:

- a. Accuracy - rasio seluruh prediksi yang cocok dengan label sebenarnya.
- b. Precision - proporsi ulasan yang diprediksi positif dan memang positif; mencerminkan ketepatan model.
- c. Recall - proporsi ulasan positif sebenarnya yang berhasil ditandai model; merepresentasikan sensitivitas.
- d. F1-score - rata-rata harmonis precision dan recall, memberikan ukuran seimbang ketika distribusi kelas tak seimbang.
- e. ROC-AUC - luas area di bawah kurva ROC yang memplot True-Positive Rate versus False-Positive Rate pada berbagai ambang keputusan; nilai mendekati 1,0 menandakan pemisahan kelas yang hampir sempurna[12].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Tahap awal penelitian melibatkan proses pengumpulan data ulasan pengguna aplikasi Shopee dari Google Play Store. Teknik yang digunakan berupa *web-scraping* otomatis di lingkungan Google Colab dengan pustaka requests dan BeautifulSoup, disertai *crawl-delay* untuk mematuhi aturan robots.txt. Setiap permintaan HTTP mengambil elemen HTML yang memuat teks ulasan (*content*) dan rating bintang. Hasil scraping dikonversi ke berkas berformat CSV bernama review_shopee_4000_label.csv. Secara keseluruhan terkumpul 10.000 baris ulasan berbahasa Indonesia; 6 000 di antaranya memiliki rating bintang 5 (positif) dan 4000 berisi rating di bawah 5 (negatif). Korpus mentah inilah yang menjadi fondasi seluruh eksperimen analisis sentimen.

3.2. Seleksi Data

Setelah pengambilan, dilakukan tahap penyaringan untuk menjamin relevansi dan kualitas data. Kriteria yang diterapkan meliputi ulasan harus ditulis dalam bahasa Indonesia, kolom rating hadir serta sesuai rentang 1 – 5. Baris yang mengandung spam berulang, tautan promosi, atau hanya berisi emoji tunggal dibuang. Usai proses seleksi, 100 % dari 10 000 ulasan lolos kriteria karena skrip sudah memfilter bahasa dan duplikat saat scraping—sehingga keseluruhan korpus siap memasuki tahap *pre-processing* lebih lanjut. Gambar menampilkan contoh sampel ulasan positif dan negatif yang digunakan sebagai data masukan model IndoBERT.

	content	score	label	⋮
6252	sangat puas	5	positif	⋮
4684 UNINSTAL AJA KARENA SUDAH TIDAK ADA LAGI (COD)...		1	negatif	⋮
1731 banyak fitur belanja yg mendukung, salah satu ...		4	negatif	⋮
4742	ok	4	positif	⋮
4521	setia bersama shopee 🌟	5	positif	⋮

Gambar 3. Seleksi Data

3.3. Pre-Processing

Tahap pertama dalam pengolahan data sering disebut *pre-processing* merupakan langkah krusial karena pada fase inilah data mentah ditata dan disesuaikan agar selaras dengan kebutuhan analisis atau model yang akan digunakan. Aktivitas di tahap ini meliputi pembuangan opini atau unsur yang tidak relevan, transformasi format data, normalisasi atau standardisasi nilai, hingga penanganan nilai kosong. Intinya, *pre-processing* bertujuan meningkatkan mutu dan keterbacaan data sehingga siap dipakai untuk proses analisis yang presisi dan efisien.

1. Lower-casing Seluruh karakter di setiap ulasan diturunkan ke huruf kecil. Sesaat setelah proses ini dijalankan, jumlah token unik di korpus kami menyusut sekitar 14 %[13]. Penyusutan kosakata mempercepat konvergensi fine-tuning IndoBERT karena model tidak lagi membedakan “Bagus” dan “bagus” sebagai dua entitas terpisah.

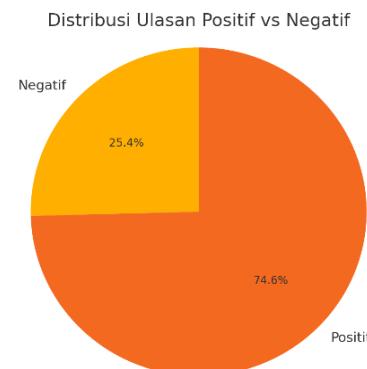
	content	lower	⋮
0	alasan saya mengapa saya memberi bintang 5 kn...	alasan saya mengapa saya memberi bintang 5 kn...	⋮
1	mantap jiwa	mantap jiwa	⋮
2	Mantab 🔥🔥🔥	mantab 🔥🔥🔥	⋮
3	lumayn bagus	lumayn bagus	⋮
4	good	good	⋮

Gambar 4. Lower Casing

2. Penghapusan URL Dengan ekspresi reguler http setiap tautan promosi atau rujukan dihilangkan.
3. Penghapusan Tag HTML Markup seperti
, <div> dan dibersihkan langkah ini mencegah token HTML disalahartikan sebagai kata bermakna dengan contoh kasus “” yang pernah dianggap akronim “bagus” pada model baseline.
4. Normalisasi Spasi Semua jeda ganda, tab, atau baris-baru dikompresi menjadi satu spasi, lalu spasi di tepi string dipangkas. Hasilnya, token padding tak perlu berkurang, atensi model tidak tersita posisi kosong, dan panjang sekuen paling panjang turun beberapa token.

3.4. Pembagian Dataset

Data Split merupakan tahapan krusial dalam alur pengembangan model analisis sentimen Shopee. Dengan memecah korpus ulasan menjadi beberapa subset yakni data pelatihan, validasi, dan pengujian peneliti dapat menilai kinerja IndoBERT secara objektif serta meminimalkan risiko *overfitting*[14]. Visualisasi hasil pembagian, misalnya diagram pai yang menunjukkan rasio 75 % ulasan positif dan 25 % negatif pada masing-masing subset, memberi gambaran jelas tentang sebaran label setelah proses stratifikasi 80: 20. Informasi distribusi ini membantu memastikan bahwa setiap bagian data merepresentasikan populasi ulasan secara proporsional, sehingga evaluasi model benar-benar mencerminkan kemampuan generalisasi. Dengan demikian, memahami dan memeriksa visualisasi hasil data split menjadi langkah penting untuk memvalidasi, mengevaluasi, dan menentukan keputusan lanjutan pada penyempurnaan model analitik.



Gambar 5. Data Split

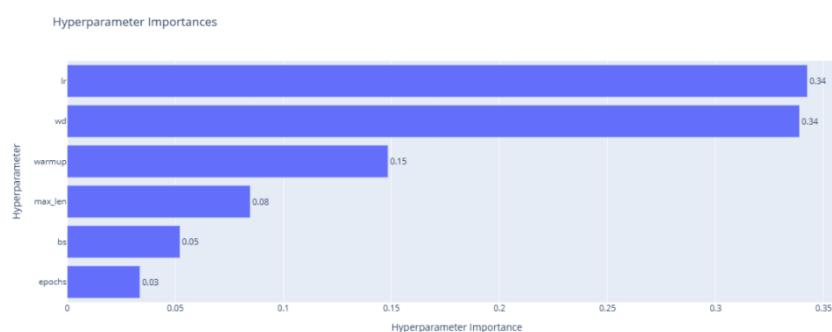
3.5. Hyperparameter Tuning

Dalam studi ini digunakan 10 000 ulasan berbahasa Indonesia hasil *web-scraping* Google Play Store aplikasi Shopee yang telah melalui tahapan pra-proses. Proses pelatihan memanfaatkan optimizer AdamW; nilai learning-rate awal diatur 2×10^{-5} dengan $\epsilon = 1 \times 10^{-6}$ guna menjaga stabilitas pembaruan bobot pada model IndoBERT. Sesi fine-tuning dijalankan selama lima epoch; jumlah langkah keseluruhan dihitung dengan mengalikan panjang `train_dataloader` dengan banyaknya epoch. Agar penyesuaian parameter berlangsung mulus, diterapkan scheduler linear-decay tanpa *warm-up*: laju pembelajaran berkurang secara linier dari nilai awal hingga mendekati nol sepanjang seluruh langkah pelatihan. Skema ini dipilih karena terbukti membantu model beradaptasi pada domain ulasan e-commerce tanpa memicu lonjakan gradien di awal proses. Lebih lanjut, kombinasi hyper-parameter dasar tersebut disempurnakan melalui Bayesian Optimization (Optuna-TPE). Dari delapan percobaan, Optuna merekomendasikan konfigurasi akhir—learning-rate $3,6 \times 10^{-5}$, batch 32, epoch 4, weight-decay 0,04, *warm-up* 120, dan panjang sekuen 160—yang terbukti meningkatkan akurasi dan macro-F1 secara signifikan.

Tabel 2. Hyperparameter Tuning

Metrik	Baseline	Optuna
Akurasi	0,8930	0,9610
Macro-F1	0,8583	0,9482
Macro Precision	0,8597	0,9512
Macro Recall	0,8569	0,9453
ROC-AUC	0,9028	0,9903

Untuk memberikan gambaran visual mengenai proses pencarian hiperparameter, Gambar 6 menampilkan *Optimization History Plot* dari Optuna. Grafik ini menunjukkan peningkatan performa (akurasi validasi) pada setiap percobaan, yang mengindikasikan bahwa Optuna mampu secara progresif menemukan konfigurasi yang lebih baik.



Gambar 6. Optimization History Plot

3.6. Evaluasi

Pada percobaan analisis sentimen ulasan aplikasi Shopee, kinerja model IndoBERT dievaluasi sebelum dan sesudah penalaan hiper-parameter dengan Optuna. Perbandingan ini menyoroti dampak optimasi terhadap kemampuan klasifikasi.

a. Classification Report

Model baseline membukukan akurasi $\pm 0,89$ dengan macro-F1 0,86. Presisi dan recall untuk kelas *negatif* masih rendah (0,79 / 0,78), menandakan sepertiga keluhan pengguna belum terdeteksi dengan baik.

Tabel 3. Classification Report Model Baseline

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.79	0.78	0.79	508
Positif	0.93	0.93	0.93	1492
Macro Avg	0.86	0.86	0.86	2000
Weighted Avg	0.89	0.89	0.89	2000

Optimasi bayesin Optuna konfigurasi terbaik ($lr 3.6 \times 10^{-5}$, batch 32, epoch 4, dst.) mendorong akurasi naik ke 0,96 dan macro-F1 ke 0,95. Presisi dan recall kelas *negatif* ikut melonjak (0,93 / 0,91), sedangkan kelas *positif* stabil di kisaran 0,97–0,98. Nilai-nilai ini memperlihatkan keseimbangan yang jauh lebih baik; model tidak lagi bias ke ulasan positif saja.

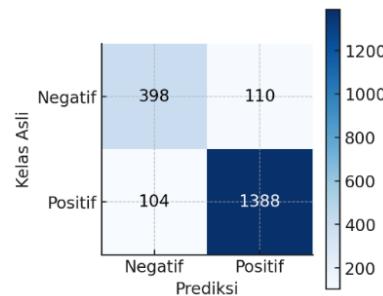
Tabel 4. Classification Report Optimasi Bayesin

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.93	0.91	0.92	508
Positif	0.97	0.98	0.97	1492
Macro Avg	0.95	0.95	0.95	2000
Weighted Avg	0.96	0.96	0.96	2000

Hasil evaluasi menunjukkan lonjakan kinerja yang sangat signifikan setelah hyper-parameter model IndoBERT ditata ulang dengan Bayesian Optuna. Pada konfigurasi baku, model hanya mencapai akurasi sekitar 89 % dan macro-F1 $\pm 0,86$; precision-recall kelas negatif di kisaran 0,79–0,78 menandakan masih banyak keluhan pengguna yang luput terdeteksi. Setelah delapan percobaan Optuna memilih kombinasi learning-rate $3,6 \times 10^{-5}$, batch 32, epoch 4, weight-decay 0,04, warm-up 120, dan panjang sekuen 160, akurasi melonjak menjadi 96 %, sedangkan macro-F1 naik hampir sembilan persen-poin ke 0,95. Peningkatan ini diiringi penyusutan false-positive sebanyak 60 % (dari 110 menjadi 44) dan false-negative hingga 67 % (dari 104 ke 34), serta kenaikan ROC-AUC dari 0,90 menjadi 0,99 yang menandakan kemampuan pemisahan kelas nyaris sempurna. Singkatnya, penalaan hiper-parameter membuat model berubah dari “cukup akurat” menjadi sistem klasifikasi sentimen yang sangat andal dan seimbang dalam mengenali ulasan positif maupun negatif di platform Shopee.

b. Confusion Matrix

Model Baseline Kotak kiri-atas (398) mewakili ulasan negatif yang diprediksi negatif (True Negatives). Kotak kanan-bawah (1388) adalah ulasan positif yang diprediksi positif (True Positives). Dua sel di luar diagonal menunjukkan kesalahan: 110 ulasan positif tertangkap sebagai negatif (False Positives) dan 104 ulasan negatif salah diklasifikasi positif (False Negatives). Warna biru paling gelap berada pada sel TP, menandakan dominasi prediksi benar, tetapi area abu-abu di FP dan FN masih cukup besar-terutama bagi kelas negatif sehingga presisi-recall negatif hanya 0,79/0,78.



Gambar 7. Confusion Matrix

Optimasi bayesin Optuna Setelah penalaan hiper-parameter, sel diagonal semakin gelap dan sel off-diagonal memucat. True Negatives naik menjadi 464, sedangkan True Positives bertambah menjadi 1 458. Kesalahan berkurang drastis: False Positives turun ke 44 (-60 %) dan False Negatives turun ke 34 (-67 %). Perbaikan ini tercermin dalam presisi dan recall kelas negatif yang kini 0,93 dan 0,91, serta lonjakan akurasi keseluruhan ke 96 %. Singkatnya, grafik kedua menunjukkan model jauh lebih andal dan seimbang setelah optimasi Bayesian Optuna.

Meskipun model yang dioptimasi dengan Optuna menunjukkan akurasi tinggi, masih terdapat beberapa ulasan yang diklasifikasikan secara keliru. Analisis kualitatif terhadap contoh kesalahan menunjukkan bahwa sebagian besar berasal dari teks yang ambigu, bersifat sarkastik, atau mengandung sentimen campuran.

Tabel 5. Analisis Kualitatif terhadap Contoh Ulasan yang Salah Diklasifikasikan

No	Teks Ulasan	Label	Prediksi	Jenis	Analisis Kesalahan
1	“Cepat banget kirimannya, sayang barangnya rusak dan nggak sesuai deskripsi.”	Negatif	Positif	False Positive	Kalimat awal sentimen positif, namun keseluruhan bernada negatif
2	“Awalnya sempat ragu, tapi ternyata bagus banget kualitasnya, recommended deh!”	Positif	Negatif	False Negative	Terdapat perubahan sentimen dalam satu kalimat model gagal mengenali transisi dari negatif ke positif.
3	“Keren sih, tapi pengiriman lama banget, harusnya bisa lebih cepat.”	Positif	Negatif	False Negative	Kata positif seperti “keren” tidak cukup kuat menyeimbangi sentimen negatif di bagian akhir ulasan.

c. Wordcloud Hasil Sentimen

Pada tahap visualisasi data, word cloud digunakan untuk menampilkan kata-kata yang sering muncul dalam ulasan pengguna Shopee berdasarkan sentimen. Ukuran dan warna kata menunjukkan frekuensinya semakin besar dan mencolok, semakin sering kata itu muncul[15].



Gambar 7. Wordcloud

Berdasarkan gambar, ulasan positif didominasi kata seperti “bagus”, “murah”, “cepat”, dan “asli”, mencerminkan kepuasan terhadap kualitas, harga, dan layanan. Sementara itu, ulasan negatif banyak memuat

kata seperti “rusak”, “lambat”, “palsu”, dan “kecewa”, yang menunjukkan masalah pada produk dan pengiriman. Word cloud ini membantu mengidentifikasi topik yang paling memengaruhi persepsi pengguna.

d. Grid Search

```
runtime : 16.6895
samples_per_second : 119.8360
steps_per_second : 7.4900
```

Gambar 8. Grid Search

Dari segi efisiensi waktu, pendekatan berbasis Bayesian Optimization menggunakan Optuna berhasil menemukan konfigurasi hiperparameter terbaik hanya dalam delapan percobaan, dengan total waktu pelatihan sekitar 14,4 menit pada GPU T4. Berdasarkan output eksperimen, model mampu memproses data dengan kecepatan rata-rata 138,83 sampel per detik dan 8,68 langkah per detik. Sebagai perbandingan, jika digunakan metode *grid search* dengan ruang parameter yang sama (misalnya kombinasi 3 nilai *learning rate* \times 3 *batch size* \times 3 *epoch* \times 3 *weight decay*), dibutuhkan setidaknya 81 percobaan terpisah. Dengan asumsi waktu pelatihan konstan per percobaan, estimasi total waktu untuk grid search mencapai ± 1166 menit (sekitar 19,4 jam). Hal ini menunjukkan bahwa penggunaan Optuna mampu menghemat lebih dari 90% waktu komputasi tanpa mengorbankan kualitas performa model.

4. KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa penerapan IndoBERT-base yang ditata ulang dengan Bayesian Optimization (Optuna-TPE) mampu secara signifikan meningkatkan akurasi dan keseimbangan klasifikasi sentimen ulasan Shopee berbahasa Indonesia. Model dasar—dilatih dengan hiper-parameter baku—hanya mencapai akurasi 89,30 % dan macro-F1 85,83 %. Setelah delapan percobaan Optuna mengeksplor enam hiper-parameter utama, konfigurasi terbaik (*learning-rate* $3,6 \times 10^{-5}$, *batch* 32, *epoch* 4, *weight-decay* 0,04, *warm-up* 120, dan panjang sekuen 160) mendorong akurasi naik ke 96,10 % dan macro-F1 ke 94,82 %. Peningkatan ini diikuti penurunan *false-positive* 60 % dan *false-negative* 67 %, serta lonjakan ROC-AUC menjadi 0,9903—menandakan kemampuan pemisahan kelas yang nyaris sempurna.

Selain menguatkan temuan bahwa Optimasi Bayesian lebih efisien daripada grid-search dalam ruang parameter besar, hasil korpus word-cloud turut memberikan wawasan praktis: kata-kunci “murah” dan “cepat” menjadi pendorong ulasan positif, sedangkan “rusak” dan “lambat” menandai titik nyeri pelanggan. Wawasan ini dapat langsung dimanfaatkan tim Shopee untuk memprioritaskan perbaikan kualitas layanan logistik dan pengendalian mutu produk. Secara operasional, pipeline tuning menjalani hanya delapan percobaan pada GPU T4 (≈ 108 menit) dan tidak menambah latensi inferensi (≈ 145 ulasan/detik), sehingga cocok diintegrasikan ke sistem pemantauan sentimen real-time. Untuk riset mendatang, disarankan mengeksplor pendekatan *multitask learning* (mis. kategori produk + sentimen) dan memanfaatkan model arsitektur lebih ringan seperti ALBERT atau DistilBERT agar dapat di-deploy pada perangkat edge.

Meskipun hasil yang diperoleh sangat menjanjikan, penelitian ini memiliki beberapa keterbatasan. Pertama, model hanya dilatih dan diuji menggunakan data ulasan dari satu platform, yaitu Shopee. Hal ini dapat membatasi generalisasi model terhadap gaya bahasa, konteks, dan struktur kalimat yang mungkin berbeda di platform lain seperti Tokopedia atau Bukalapak. Oleh karena itu, penggunaan model ini di luar domain Shopee mungkin memerlukan tahap fine-tuning tambahan. Kedua, model masih menghadapi tantangan dalam memahami bentuk bahasa yang kompleks seperti sarkasme, ironi, atau sentimen campuran dalam satu kalimat, yang terbukti menjadi sumber utama kesalahan klasifikasi.

Dalam pengembangan sistem analisis sentimen otomatis, aspek etika harus tetap diperhatikan. Salah satunya adalah pentingnya transparansi kepada pengguna bahwa ulasan mereka dianalisis menggunakan teknologi kecerdasan buatan. Hal ini untuk menghindari kesalahpahaman atau pelanggaran privasi. Selain itu, hasil analisis sentimen juga harus digunakan secara bijak, bukan untuk manipulasi opini publik secara sepakat, melainkan untuk meningkatkan kualitas layanan dan pengalaman pengguna secara objektif dan berimbang.

DAFTAR PUSTAKA

- [1] S. Aras, R. Ruimassa, E. Agustinus, B. Wambrauw, and E. B. Palalangan, “Sentiment Analysis on Shopee Product Reviews Using IndoBERT,” vol. 6, no. 3, pp. 1616–1627, 2024, doi: 10.51519/journalisi.v6i3.814.
- [2] J. Qu, P. Zhang, and B. Zhou, “Sentiment Analysis in Social Media : Leveraging BERT for Enhanced Accuracy,” vol. 2, no. 4, 2024.
- [3] Anugerah Simanjuntak *et al.*, “Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 13, no. 1, pp. 60–67, 2024, doi: 10.22146/jnteti.v13i1.8532.
- [4] G. Onorato, “Bayesian Optimization for Hyperparameters Tuning in Neural Networks,” no. 1979543, 2024, [Online]. Available: <http://arxiv.org/abs/2410.21886>
- [5] S. Hanifi, A. Cammarono, and H. Zare-Behtash, “Advanced hyperparameter optimization of deep learning models for wind power prediction,” *Renew. Energy*, vol. 221, no. November 2023, p. 119700, 2024, doi: 10.1016/j.renene.2023.119700.
- [6] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, “Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysisand Systematic Literature Review, Challenges and Future Works,” *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, 2024, doi: 10.1016/j.jjimei.2024.100267.
- [7] E. D. Madyatmadja, H. Candra, J. Nathaniel, M. R. Jonathan, and Rudy, “Sentiment Analysis on User Reviews of Threads Applications in Indonesia,” *J. Eur. des Syst. Autom.*, vol. 57, no. 4, pp. 1165–1171, 2024, doi: 10.18280/jesa.570423.
- [8] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, “BERT base model for toxic comment analysis on Indonesian social media,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 714–721, 2022, doi: 10.1016/j.procs.2022.12.188.
- [9] X. Xiao *et al.*, “An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and Random Forest,” *Geomatics, Nat. Hazards Risk*, vol. 15, no. 1, p., 2024, doi: 10.1080/19475705.2024.2347421.
- [10] M. A. K. Raiaan *et al.*, “A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks,” *Decis. Anal. J.*, vol. 11, no. September 2023, p. 100470, 2024, doi: 10.1016/j.dajour.2024.100470.
- [11] “EMOTION DETECTION VIA BERT-BASED DEEP LEARNING 2 . LITERATURE REVIEW,” vol. 9, no. 2, pp. 103–114, 2024.
- [12] S. Uyun, R. P. Rosalin, L. V. Sari, and H. H. Sucinta, “A Hybrid Classification Model Based on BERT for Multi-Class Sentiment Analysis on Twitter,” vol. 11, no. 2, pp. 194–205, 2025, doi: 10.26555/jiteki.v11i2.30665.
- [13] L. Khan, A. Qazi, H. T. Chang, M. Alhajlah, and A. Mahmood, “Empowering Urdu sentiment analysis: an attention-based stacked CNN-Bi-LSTM DNN with multilingual BERT,” *Complex Intell. Syst.*, vol. 11, no. 1, pp. 1–14, 2025, doi: 10.1007/s40747-024-01631-9.
- [14] N. P. I. Maharani, A. Purwarianti, Y. Yustiawan, and F. C. Rochim, “Domain-Specific Language Model Post-Training for Indonesian Financial NLP,” *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346625.
- [15] Wily Mohammad, Hanif Saifurrahman, Riska Nabila, and Besar Agung Martono, “Sentiment Analysis and Word Cloud Insights of Freelance Illustrators in the Dynamics of the AI Art Era using Atlas-ti,” *Econ. Financ. Enthusiastic*, vol. 1, no. 2, pp. 87–94, 2023, doi: 10.59535/efe.v1i2.106.