

Klasifikasi Sentimen Ulasan Produk pada Platform E-Commerce di Indonesia dengan Menggunakan Model Pre-Trained IndoBERT

Bayu Puspito Aji*, Christian Sri Kusuma Aditya

Fakultas Teknik, Program Studi Informatika, Universitas Muhammadiyah Malang, Kota Malang, Indonesia

Email: ^{1,*}bayupspto_aji@webmail.umm.ac.id, ²christianskaditya@umm.ac.id

Email Penulis Korespondensi: bayupspto_aji@webmail.umm.ac.id

Submitted: 11/02/2025; Accepted: 05/03/2025; Published: 07/03/2025

Abstrak—Dalam era digital saat ini, analisis sentimen terhadap ulasan produk di platform *e-commerce* menjadi semakin penting, terutama di Tokopedia, salah satu marketplace terbesar di Indonesia. Tokopedia menyediakan fasilitas bagi pengguna untuk memberikan ulasan setelah melakukan transaksi, yang berperan penting dalam membantu pelaku bisnis memahami persepsi pelanggan terhadap produk. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan produk di Tokopedia menggunakan model *IndoBERT* serta mengevaluasi performanya dibandingkan dengan metode berbasis *LSTM* yang dikombinasikan dengan *embedding FastText, Glove, dan Word2Vec*. Model *LSTM-FastText* dalam penelitian sebelumnya berhasil mencapai akurasi tertinggi sebesar 85,08%. Pada penelitian ini, klasifikasi sentimen ulasan produk di Tokopedia dilakukan dengan jumlah data sebanyak 5400 dan proses klasifikasi sentimen dibagi menjadi dua kategori, yaitu *positif* dan *negatif*, dengan pembagian dataset menjadi tiga kelompok: pelatihan, validasi, dan pengujian. Kontribusi dalam penelitian ini adalah mengeksplorasi efektivitas kinerja model *IndoBERT* dibandingkan dengan metode sebelumnya yang mengimplementasikan model *LSTM* bersama *embedding FastText, Glove, dan Word2Vec*. Berdasarkan hasil penelitian, model *IndoBERT* mencapai *accuracy* sebesar 97%, dengan nilai *F1-score* yang sama untuk kedua kategori sentimen sebesar 97%. Dirancang khusus dengan *pre-training* pada korpus besar berbahasa Indonesia, *IndoBERT* mampu memahami konteks teks dengan lebih baik dibandingkan dengan model *LSTM* yang digunakan pada penelitian sebelumnya. Hal ini memungkinkan *IndoBERT* untuk menghasilkan akurasi yang lebih tinggi, karena dapat memahami ulasan produk dalam bahasa Indonesia dengan lebih efektif.

Kata Kunci: Analisis Sentimen; IndoBERT; Ulasan Produk; Deep Learning; E-Commerce

Abstract—In today's digital era, sentiment analysis of product reviews on e-commerce platforms is becoming increasingly important, especially on Tokopedia, one of the largest marketplaces in Indonesia. Tokopedia provides facilities for users to leave reviews after making transactions, which play an important role in helping businesses understand customer perceptions of products. This research aims to classify the sentiment of product reviews on Tokopedia using the IndoBERT model and evaluate its performance compared to LSTM-based methods combined with FastText, Glove, and Word2Vec embedding. The LSTM-FastText model in previous research achieved the highest accuracy of 85.08%. In this study, the sentiment classification of product reviews on Tokopedia was carried out with a total of 5400 data and the sentiment classification process was divided into two categories, namely positive and negative, with the division of the dataset into three groups: training, validation, and testing. The contribution in this research is to explore the effectiveness of the IndoBERT model performance compared to previous methods that implement the LSTM model with FastText, Glove, and Word2Vec embedding. Based on the research results, the IndoBERT model achieved an accuracy of 97%, with the same F1-score value for both sentiment categories of 97%. Specifically designed with pre-training on a large Indonesian corpus, IndoBERT is able to understand the context of the text better than the LSTM model used in previous studies. This allows IndoBERT to produce higher accuracy, as it can understand product reviews in Indonesian more effectively.

Keywords: Sentiment Analysis; IndoBERT; Product Reviews; Deep Learning; E-Commerce

1. PENDAHULUAN

Marketplace adalah platform *e-commerce* yang berfungsi sebagai perantara antara penjual dan pembeli, serta menjadi tempat untuk menjalankan aktivitas bisnis. Dalam konteks analisis sentimen, ulasan pengguna di platform *e-commerce* dapat dianalisis untuk memahami persepsi konsumen terhadap suatu produk. Analisis tersebut memiliki peran dalam mengevaluasi tingkat kepuasan pelanggan, mengidentifikasi tren pasar, serta meninjau aspek yang dapat ditingkatkan oleh penjual atau penyedia layanan[1]. Berdasarkan data yang diperoleh dari situs iPrice, Tokopedia menduduki peringkat pertama sebagai marketplace dengan jumlah kunjungan tertinggi, mencapai 157 juta kunjungan per bulan hingga kuartal ketiga tahun 2021. Kemajuan teknologi yang pesat telah mendorong perubahan perilaku konsumen dari berbelanja secara offline ke online, termasuk di Indonesia. Hal ini terlihat dari pertumbuhan *e-commerce* yang mengalami peningkatan signifikan. Tokopedia, sebagai salah satu platform *e-commerce* terbesar di Indonesia, menyediakan berbagai fitur yang dapat membantu konsumen dalam memilih produk yang akan dibeli. Salah satu fitur penting yang digunakan sebagai sumber informasi oleh konsumen adalah ulasan dan penilaian (*rating*) produk[2]. Mengingat volume transaksi yang besar, diperlukan sistem klasifikasi sentimen otomatis yang dapat membantu menganalisis ulasan produk dalam skala yang luas. Klasifikasi sentimen ini penting bagi penjual, karena memungkinkan mereka untuk memahami opini konsumen terhadap produk yang mereka jual, baik berupa ulasan positif maupun negatif[3].

Penelitian sebelumnya terkait analisis sentimen ulasan produk telah menggunakan teknik-teknik tradisional seperti *TextBlob, Naive Bayes & K-NN (K-Nearest Neighbor)*, namun teknik-teknik ini memiliki keterbatasan dalam memahami konteks yang kompleks dari sebuah kalimat. Pendekatan teknik-teknik tersebut menggunakan fitur berbasis kata tanpa memperhitungkan hubungan semantik dalam struktur kalimat atau variasi penggunaan kata dalam berbagai konteks. Hal ini berdampak pada tingkat akurasi dalam analisis sentimen, khususnya dalam bahasa Indonesia

yang memiliki beragam bentuk kata. Dari semua model tersebut, hasil penelitian menunjukkan bahwa model *K-NN* mempunyai akurasi terbaik diantara *TextBlob* dan *Naive Bayes* yaitu sebesar 75%, sedangkan metode *Naive Bayes* memiliki akurasi sebesar 69.44%, disusul oleh *TextBlob* sebagai metode dengan akurasi terendah sebesar 54.67% [4]. Penelitian serupa lainnya menggunakan *CNN* (Convolutional Neural Network) dan *SVM* (Support Vector Machine) dengan *Hybrid Optimization*. Hasilnya menunjukkan bahwa model *CNN* mencapai akurasi sebesar 87,62%, sedangkan model *SVM* dengan *Hybrid Optimization* menghasilkan akurasi yang baik sebesar 91,91% [5]. Penelitian sebelumnya mengenai klasifikasi sentimen dengan dataset yang serupa telah menggunakan metode *LSTM* (Long Short-Term Memory) bersama dengan tiga *pre-trained embedding*, yaitu *FastText*, *Glove*, dan *Word2Vec*. Model *LSTM* dengan embedding *FastText* menghasilkan akurasi tertinggi sebesar 85,08%. Berdasarkan hasil tersebut, penulis merekomendasikan penggunaan model *IndoBERT* sebagai alternatif untuk meningkatkan kinerja klasifikasi sentimen pada dataset yang serupa [6].

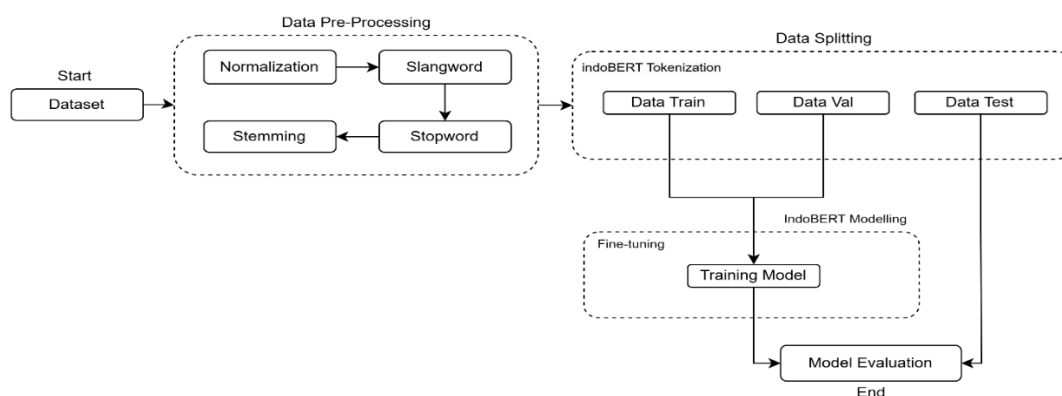
Pengenalan model berbasis *transformer* seperti *BERT* (Bidirectional Encoder Representations from Transformers), telah membawa perubahan signifikan dalam bidang pemrosesan bahasa alami (NLP). *BERT* menggunakan arsitektur *transformer* yang dapat memahami konteks teks secara dua arah secara bersamaan, yang terbukti meningkatkan akurasi dalam berbagai tugas NLP, termasuk dalam klasifikasi sentimen [7]. Penelitian sebelumnya telah menggunakan data ulasan produk di platform *e-commerce* Amazon dengan penerapan model *BERT*, *LSTM*, *SVM*, dan *Naive Bayes*. Hasilnya menunjukkan bahwa model *BERT* mencapai akurasi tertinggi, yaitu sebesar 88.80%. Temuan ini menunjukkan bahwa *BERT* mampu memberikan kinerja yang lebih unggul dalam memprediksi sentimen dibandingkan pendekatan berbasis ekstraksi fitur tradisional dan model pembelajaran mesin lainnya [8]. Model *IndoBERT* telah diterapkan dalam penelitian analisis laporan radiologi, dan menghasilkan akurasi model yang mencapai lebih dari 90%. Hasil ini menunjukkan efektivitas *IndoBERT* dalam memahami dan mengolah teks dalam bahasa Indonesia pada konteks medis [9]. Berbagai penelitian telah mengidentifikasi keunggulan model *BERT* dalam pemrosesan bahasa alami. Namun, terdapat keterbatasan dalam penggunaan model *pre-trained* yang secara khusus dirancang untuk bahasa Indonesia, terutama pada analisis sentimen ulasan di platform *e-commerce* Indonesia. Meskipun beberapa penelitian telah menerapkan model ini, penerapan yang lebih spesifik dan relevan terhadap bahasa Indonesia masih jarang ditemukan.

Berdasarkan penelitian-penelitian tersebut, teridentifikasi adanya gap dalam penggunaan model tradisional untuk klasifikasi sentimen terutama dalam memahami konteks bahasa yang kompleks. Model-model seperti *Naive Bayes*, *K-NN*, dan *TextBlob* menunjukkan performa yang kurang optimal dalam menangkap makna kalimat yang lebih mendalam, terutama dalam bahasa Indonesia. Hal ini menjadi kendala signifikan karena ulasan produk seringkali mengandung opini yang tidak hanya bersifat sederhana, tetapi juga kompleks dan kontekstual. Dengan menggunakan *IndoBERT*, penelitian ini berkontribusi pada peningkatan kinerja klasifikasi sentimen produk di Indonesia dengan memanfaatkan model berbasis *transformer* yang secara khusus dirancang untuk bahasa Indonesia, yang sebelumnya jarang diterapkan dalam konteks *e-commerce*. Penelitian ini bertujuan untuk mengatasi kekurangan yang ada dengan menerapkan model *pre-trained IndoBERT* untuk klasifikasi sentimen ulasan produk di Tokopedia dari dataset *PRDECT-ID*. Diharapkan bahwa penggunaan *IndoBERT* akan meningkatkan performa klasifikasi sentimen dibandingkan dengan model dari penelitian sebelumnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan mengikuti serangkaian langkah yang terstruktur untuk memastikan proses klasifikasi data berjalan lancar dan tujuan penelitian tercapai. Setiap tahapan dalam penelitian ini dijelaskan melalui sebuah diagram alur, yang menggambarkan langkah-langkah yang diambil secara teratur dan sistematis. Gambar 1 menunjukkan rincian alur penelitian tersebut, memberikan gambaran jelas mengenai setiap proses yang dilalui.



Gambar 1. Alur Proses Penelitian



Proses penelitian ini dimulai dari pemrosesan data hingga evaluasi model. Dataset pertama kali diproses melalui tahap *preprocessing* untuk memastikan kualitasnya, seperti pembersihan data dan penyesuaian format. Setelah itu, data dibagi menjadi tiga kelompok: data pelatihan (*training*), validasi (*validation*), dan pengujian (*testing*). Selanjutnya, model *IndoBERT* digunakan dengan memulai tahap *pre-training* dan *tokenization*, yaitu mengubah teks menjadi bentuk yang bisa dipahami oleh model. Setelah itu, dilakukan *fine-tuning*, yaitu penyesuaian lebih lanjut terhadap model agar lebih optimal berdasarkan data yang digunakan dalam penelitian ini. Tahap akhir adalah *model evaluation*, di mana kinerja model dinilai untuk melihat seberapa baik model dapat mengklasifikasikan sentimen ulasan dengan akurat. Setiap langkah dalam proses ini dirancang secara sistematis agar model yang dihasilkan mampu memprediksi sentimen ulasan produk di Tokopedia dengan hasil yang lebih tepat.

2.2 Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari *Mendeley Data* dengan nama *PRDECT-ID*, singkatan dari *Product Reviews Dataset for Emotion Classification Task*[10]. Dataset ini berisi ulasan produk dalam bahasa Indonesia yang telah diberi anotasi sentimen. Sentimen dalam dataset terbagi menjadi dua kategori, yaitu Positif dan Negatif dengan total 5400 data.

2.3 Data Preprocessing

Tahap *preprocessing data* adalah tahapan penting dalam persiapan data sebelum analisis. Data yang diperoleh dalam penelitian ini tidak siap untuk langsung digunakan, karena mengandung berbagai bentuk teks yang tidak konsisten, penggunaan bahasa gaul, serta elemen yang tidak relevan untuk analisis sentimen[11]. Oleh karena itu, proses *preprocessing* dilakukan guna meningkatkan kualitas data dan meningkatkan performa model prediksi. Proses *preprocessing* melibatkan beberapa tahapan utama, yaitu *normalization*, penggantian kata slang (*slang word*), penghapusan kata berhenti (*stopwords*), dan *stemming*. Setiap tahapan ini memiliki fungsi tertentu dalam menyederhanakan dan membersihkan data teks agar dapat diolah secara optimal oleh model. Tahapan *preprocessing* tidak hanya berfokus pada penghilangan ketidakpastian dalam teks, tetapi juga menjaga konsistensi bentuk kata, mengurangi kompleksitas data, dan meningkatkan akurasi model dalam memahami konteks sentimen dari ulasan produk yang dianalisis.

2.3.1 Normalization

Proses normalisasi teks digunakan untuk menyetarakan data ulasan pelanggan dengan menghilangkan elemen-elemen yang tidak diperlukan, seperti tanda baca berlebihan, simbol yang tidak relevan, dan pengubahan teks menjadi huruf kecil (*lowercase*)[12]. Langkah ini dilakukan untuk menyederhanakan teks sehingga memiliki struktur yang lebih rapi dan seragam, serta siap untuk analisis lebih mendalam. Tabel 1 menyajikan perbandingan antara teks asli ulasan pelanggan dan teks setelah melalui tahap normalisasi.

Tabel 1. Sampel *Normalization*

Ulasan Pelanggan	Teks <i>Normalization</i>
produk selalu fresh dan lezat, packing rapih, pengiriman cepat, dan pelayanan toko sangat ramah ?? sukses terus ya. ??	produk selalu fresh dan lezat packing rapih pengiriman cepat dan pelayanan toko sangat ramah sukses terus ya

Tabel 1 menampilkan hasil dari normalisasi yang melibatkan penghapusan tanda baca berlebihan, seperti tanda tanya ganda (??), konversi semua huruf menjadi huruf kecil (*lowercase*), serta penyederhanaan struktur kalimat. Hasil normalisasi ini menghasilkan teks yang lebih terstruktur, yang dapat digunakan untuk keperluan analisis lebih lanjut dalam penelitian atau evaluasi data.

2.3.2 Slangwords

Dalam analisis ulasan pelanggan, ditemukan adanya penggunaan *slangword* atau kata-kata tidak baku yang dapat mempengaruhi hasil pemrosesan teks secara otomatis. Oleh karena itu, diperlukan normalisasi teks untuk mengonversi kata-kata tidak baku menjadi bentuk yang lebih formal dan terstruktur. Proses ini melibatkan penggantian kata-kata *slang* dengan kata resmi yang memiliki makna sama. Hasil normalisasi teks dari ulasan pelanggan ditampilkan pada Tabel 2.

Tabel 2. Sampel *Slangwords*

Ulasan Pelanggan	Teks <i>Slangwords</i>
produk selalu fresh dan lezat packing rapih pengiriman cepat dan pelayanan toko sangat ramah sukses terus ya	produk selalu fresh dan lezat packing rapi pengiriman cepat dan pelayanan toko sangat ramah sukses terus ya

Tabel 2 menunjukkan normalisasi teks, di mana *slangword* seperti "rapih" digantikan dengan "rapi", serta penyesuaian lainnya sesuai dengan kaidah bahasa formal. Normalisasi teks ini dilakukan untuk meningkatkan akurasi

analisis teks dan memberikan struktur yang lebih seragam untuk proses lanjutan menggunakan algoritma pemrosesan bahasa alami.

2.3.3 Stopwords

Penghapusan *stopword* dilakukan untuk mengurangi kata-kata umum yang tidak memiliki peran signifikan dalam analisis teks. *Stopword* biasanya terdiri dari kata-kata penghubung, kata ganti, serta kata yang sering digunakan dalam bahasa sehari-hari tetapi tidak relevan dalam konteks analisis data[13]. Tabel 3 menyajikan contoh teks ulasan pelanggan sebelum dan sesudah proses penghapusan *stopword*, dengan tujuan untuk mempertahankan kata-kata yang memiliki makna substantif.

Tabel 3. Sampel *Stopwords*

Ulasan Pelanggan	Teks <i>Stopwords</i>
produk selalu fresh dan lezat packing rapi pengiriman cepat dan pelayanan toko sangat ramah sukses terus ya	produk fresh lezat packing rapi pengiriman cepat pelayanan toko ramah sukses terus ya

Tabel 3 menunjukkan hasil dari penghapusan *stopword*, di mana teks ulasan pelanggan mengalami reduksi jumlah kata tanpa mengubah informasi utama yang disampaikan. Kata-kata seperti "dan," dan karakter non-verbal telah dihapus, sementara kata-kata yang relevan seperti "fresh," "lezat," "pengiriman cepat," dan "sukses" tetap dipertahankan. Proses ini merupakan bagian penting dalam tahapan pra-pemrosesan teks yang bertujuan untuk meningkatkan efektivitas analisis.

2.3.4 Stemming

Stemming merupakan teknik dalam pemrosesan teks yang berfungsi untuk mengubah kata-kata menjadi bentuk dasar atau akar katanya. *Stemming* bekerja dengan mengidentifikasi afiks seperti awalan, akhiran, sisipan yang melekat pada kata dan menghapusnya untuk mendapatkan bentuk dasar dari kata tersebut[14]. Dengan demikian, kata-kata yang memiliki makna serupa dapat dikenali sebagai entitas yang sama. Tabel 4 menyajikan contoh teks ulasan pelanggan sebelum dan sesudah proses stemming diterapkan.

Tabel 4. Sampel *Stemming*

Ulasan Pelanggan	Teks <i>Stemming</i>
produk fresh lezat packing rapi pengiriman cepat pelayanan toko ramah sukses terus ya	produk fresh lezat packing rapi kirim cepat layan toko ramah sukses terus ya

Tabel 4 menunjukkan hasil *stemming* di mana kata seperti "pengiriman" diubah menjadi "kirim" dan "pelayanan" menjadi "layan." Proses ini mengubah kata-kata berimbuhan menjadi bentuk dasarnya. Dengan menggunakan *stemming*, kata-kata dengan makna yang sama tetapi dalam bentuk *morfologis* yang berbeda dapat diperlakukan sebagai satu kata dalam analisis teks.

2.4 Data Splitting

Proses *data splitting* digunakan untuk membagi dataset menjadi tiga bagian, yaitu data latih (*train*), data validasi (*validation*), dan data uji (*test*). Pembagian ini bertujuan untuk mengukur kinerja model secara terpisah dalam setiap tahap pengembangan. Data latih digunakan untuk membangun model, data validasi untuk memantau dan menyesuaikan parameter model selama pelatihan, serta data uji untuk menilai performa akhir model[15]. Distribusi data untuk setiap kategori ditampilkan pada Tabel 5.

Tabel 5. *Distribusi Data*

Jenis Data	Total Data
Data Latih	3456
Data Validasi	864
Data Uji	1080

Pada Tabel 5 terlihat bahwa data latih terdiri dari 3.456 sampel, data uji berjumlah 1.080 sampel, dan data validasi terdiri dari 864 sampel. Jumlah data yang bervariasi pada setiap kategori digunakan untuk memberikan porsi data yang berbeda-beda pada proses pelatihan, validasi, dan pengujian. Pendekatan ini bertujuan agar model dapat dilatih dengan data yang memadai, sehingga dapat mempelajari pola dengan efektif, sambil tetap memiliki cukup data untuk diuji dan divalidasi agar kinerja model dapat dievaluasi secara akurat.

2.5 IndoBERT Tokenization

Tokenization merupakan langkah penting dalam pemrosesan teks untuk model berbasis *BERT*, termasuk *IndoBERT*. Pada tahap ini, teks dipisahkan menjadi unit-unit yang lebih kecil seperti kata atau sub-kata, yang disebut token. Model *BERT* menggunakan metode *WordPiece* untuk melakukan tokenisasi, memungkinkan pemrosesan kata-kata yang

jarang atau tidak dikenali melalui pembagian menjadi sub-kata yang lebih umum. Setiap token yang dihasilkan kemudian dipetakan ke indeks bilangan bulat berdasarkan kamus yang telah dilatih sebelumnya[16].

Tabel 6. *Tokenization*

Ulasan Pelanggan	<i>Tokenization</i>	<i>Encode</i>
barang sesuai gambar payah	['[CLS]', 'barang', 'sesuai', 'gambar', 'payah', '[SEP]', '[PAD]', '[PAD]', ...] (max length)	([2, 963, 786, 1153, 11638, 3, 0, 0, 0, ...]) (max length)

Tabel 6 menampilkan contoh proses tokenisasi terhadap ulasan produk "*barang sesuai gambar payah*". Dalam proses tersebut, terdapat token khusus seperti [CLS] untuk menandai awal masukan, [SEP] untuk menandai akhir, dan [PAD] untuk menyelaraskan panjang token dengan panjang maksimum, yaitu 128 token, di mana urutan token dihasilkan dalam format yang kompatibel dengan model *BERT*. Setelah tokenisasi, setiap token dikonversi menjadi representasi numerik menggunakan indeks dalam kamus *IndoBERT* yang telah dilatih sebelumnya. Masing-masing kata atau sub-kata dalam ulasan produk dipetakan ke bilangan bulat yang sesuai dalam kamus tersebut. Selain itu, token [PAD] digunakan untuk memastikan semua masukan memiliki panjang yang sama agar dapat diproses dalam batch[17].

Tahap *encoding* ini penting karena model *BERT* memerlukan input dalam bentuk bilangan bulat untuk diproses lebih lanjut. Melalui *tokenization* dan *encoding*, teks diubah ke dalam format yang dapat diproses oleh model untuk tugas-tugas seperti klasifikasi sentimen[18]. Proses ini memungkinkan model untuk menangani teks dari berbagai panjang dengan efisiensi dan konsistensi.

2.6 IndoBERT Modelling

IndoBERT merupakan model berbasis arsitektur Transformer yang dilatih menggunakan kerangka kerja *Huggingface* dengan konfigurasi default *IndoBERT-Base (uncased)*. Proses pelatihan *IndoBERT* melibatkan lebih dari 220 juta kata yang diambil dari tiga sumber utama: Wikipedia (74 juta kata), berita dari Kompas, Tempo, dan Liputan 6 (55 juta kata), serta korpus web berbahasa Indonesia (90 juta kata). Model ini terdiri dari 12 lapisan tersembunyi dengan ukuran vektor 768 dimensi per lapisan, 12 attention head, dan satu lapisan feed-forward tersembunyi dengan ukuran 3.072d[19]. Pada tahap pemodelan *IndoBERT*, pemilihan parameter pelatihan dilakukan untuk mengoptimalkan kinerja model. Setiap parameter berperan penting dalam mengarahkan proses pelatihan agar berjalan secara efisien dan sesuai dengan tujuan penelitian. Parameter-parameter seperti *optimizer*, *Learning Rate*, *Batch Size*, *Max Length*, dan jumlah *epoch* ditetapkan menggunakan pendekatan trial and error atau mencoba berbagai konfigurasi secara berulang untuk menemukan kombinasi parameter yang paling optimal[20]. Setelah kombinasi parameter ditemukan, dilakukan fine-tuning pada model pre-trained *IndoBERT*. Fine-tuning adalah proses penyesuaian lebih lanjut terhadap model pre-trained dengan tujuan untuk meningkatkan kinerja model, sehingga model dapat memberikan prediksi yang lebih akurat dalam analisis sentimen[21]. Rincian parameter yang digunakan dalam penelitian ini ditunjukkan pada Tabel 7.

Tabel 7. *Hyperparameter*

Parameter	Nilai
Optimizer	Adam
Learning Rate	1e-7
Batch Size	16
Max Length	128
Epoch	41

Tabel 7 menunjukkan penggunaan *optimizer* Adam, yang dipilih karena kemampuannya dalam mengatasi masalah *sparse gradient* selama pelatihan model. Laju pembelajaran ditetapkan pada 1e-7, yang diatur untuk memastikan kestabilan selama proses pembelajaran model berukuran besar seperti *BERT*. Ukuran *batch* sebesar 16 dipilih untuk menyesuaikan dengan kapasitas memori GPU yang tersedia, serta untuk mempertahankan stabilitas pelatihan. Panjang maksimal input ditetapkan sebesar 128 token, yang sesuai untuk teks dengan panjang sedang. Jumlah *epoch* sebanyak 41 digunakan untuk memberikan cukup waktu bagi model untuk belajar dari data yang tersedia tanpa menyebabkan *overfitting*. Pemilihan parameter ini dilakukan dengan tujuan untuk mendapatkan model yang memiliki generalisasi yang baik dan kinerja yang optimal pada data uji[22].

2.7 Model Evaluation

Tahap evaluasi model bertujuan untuk mengukur kinerja model setelah proses pelatihan selesai. Evaluasi ini dilakukan menggunakan data uji yang tidak digunakan dalam proses pelatihan, untuk memberikan gambaran tentang kemampuan model dalam memprediksi data baru[23]. Beberapa metrik digunakan dalam evaluasi ini, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Setiap metrik memberikan perspektif berbeda mengenai kinerja model, mulai dari ketepatan prediksi hingga kemampuan model dalam menangani ketidakseimbangan kelas[24]. Secara matematis, dirumuskan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = 2x \frac{precision \times recall}{precision + recall} \tag{4}$$

Pada rumus di atas, TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*) merupakan indikator yang digunakan untuk menilai keakuratan model dalam mengklasifikasikan data sesuai dengan label yang sebenarnya. Pemilihan metrik evaluasi bergantung pada tujuan dan karakteristik data. *Accuracy* memberikan gambaran umum mengenai kinerja model pada seluruh data, sedangkan *precision*, *recall*, dan *F1-score* digunakan untuk menganalisis kinerja model pada masing-masing kelas, khususnya ketika data tidak seimbang[25]. Penggunaan berbagai metrik ini memberikan pandangan yang komprehensif terhadap performa model dalam skenario yang berbeda.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Dataset ini terdiri dari 5.400 ulasan produk dalam bahasa Indonesia yang telah diberi anotasi sentimen ke dalam dua kategori, yaitu positif dan negatif dan memiliki 11 fitur, yaitu *Category*, *Product Name*, *Location*, *Price*, *Overall Rating*, *Number Sold*, *Total Review*, *Customer Rating*, *Customer Review*, *Sentiment*, dan *Happy*. Dalam penelitian ini, fitur *Customer Review* digunakan sebagai data untuk analisis, sementara fitur *Sentiment* berfungsi sebagai label atau target dalam proses prediksi. Distribusi data berdasarkan kategori sentimen disajikan pada Tabel 8.

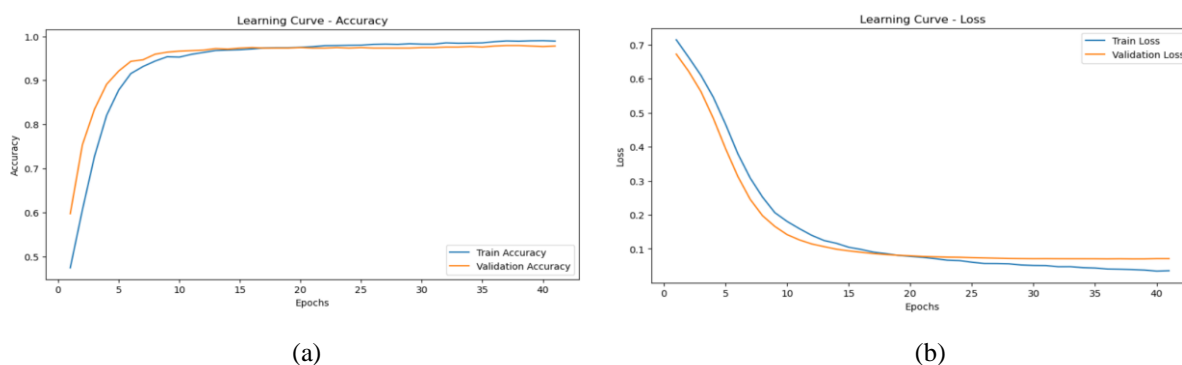
Tabel 8. Distribusi Sentimen

Sentiment	Total Data
Negatif	2821
Positif	2579

Berdasarkan Tabel 8, jumlah data dengan sentimen *Negatif* adalah 2821, sedangkan data dengan sentimen *Positif* berjumlah 2579. Distribusi data yang relatif seimbang antara kedua kategori sentimen ini memberikan peluang yang memadai untuk melatih model *deep learning* secara optimal pada kedua jenis sentimen.

3.2 Performa Model

Evaluasi performa model dapat dilakukan melalui visualisasi metrik-metrik kinerja seperti *accuracy* dan *loss*, yang memberikan informasi mengenai kemampuan model dalam mempelajari pola dari data pada tahap pelatihan dan validasi. Metrik ini penting untuk mengidentifikasi potensi masalah seperti *overfitting* dan *underfitting*. Visualisasi *learning curve* merupakan alat yang berguna dalam mengamati perkembangan performa model dari waktu ke waktu, memastikan model bekerja dengan baik pada data pelatihan dan tetap mampu menggeneralisasi dengan baik pada data validasi. Visualisasi *learning curve* memberikan informasi penting terkait perkembangan *accuracy* dan *loss*, baik pada data pelatihan maupun pada data validasi. Pada model yang ideal, performa pada data pelatihan dan validasi harus memiliki kecocokan yang baik, dengan kurva *accuracy* yang meningkat secara konsisten dan kurva *loss* yang menurun seiring bertambahnya jumlah *epoch*. Kondisi ini menandakan bahwa model mampu belajar dengan baik dari data tanpa mengalami masalah *overfitting* dan *underfitting*.



Gambar 2. (a) *Learning Curve – Accuracy* dan (b) *Learning Curve – Loss*

Gambar 2(a) menunjukkan kurva *accuracy* pelatihan dan validasi yang memperlihatkan pola yang serupa. Pada tahap awal pelatihan, terjadi peningkatan akurasi yang cepat, menandakan bahwa model mampu mengenali pola dasar

dari data secara efektif. Setelah sekitar 10 epoch, kurva *accuracy* cenderung stabil, menandakan bahwa model telah mencapai performa optimal dalam mengenali pola dari data. Kesesuaian antara kurva *accuracy* data pelatihan dan validasi menunjukkan bahwa model memiliki kemampuan untuk menggeneralisasi pola yang dipelajari dengan baik ke data baru. Tidak adanya perbedaan signifikan antara kedua kurva ini menunjukkan bahwa model tidak mengalami *overfitting*, dan model dapat mengklasifikasikan ulasan produk dengan tingkat akurasi yang mendekati optimal, yaitu sekitar 97%.

Gambar 2(b) memperlihatkan penurunan loss yang signifikan pada data pelatihan dan validasi di awal pelatihan. Kurva yang menurun tajam mengindikasikan bahwa model cepat memperbaiki kesalahan prediksi dan menyesuaikan parameter secara efektif. Setelah 10 epoch, kurva loss mulai stabil, yang menunjukkan bahwa model telah mencapai keseimbangan dalam mempelajari data pelatihan dan mempertahankan generalisasi yang baik pada data validasi. Kesesuaian antara kurva loss pada data pelatihan dan validasi menunjukkan bahwa model tidak mengalami *overfitting* dan *underfitting*. Setelah sekitar 10 epoch, kurva stabil, mengindikasikan bahwa model telah mencapai keseimbangan antara belajar dari data pelatihan dan mempertahankan generalisasi yang baik pada data validasi. Nilai loss 0.03 menunjukkan bahwa model berhasil mengurangi kesalahan prediksi secara keseluruhan.

3.3 Evaluasi Model

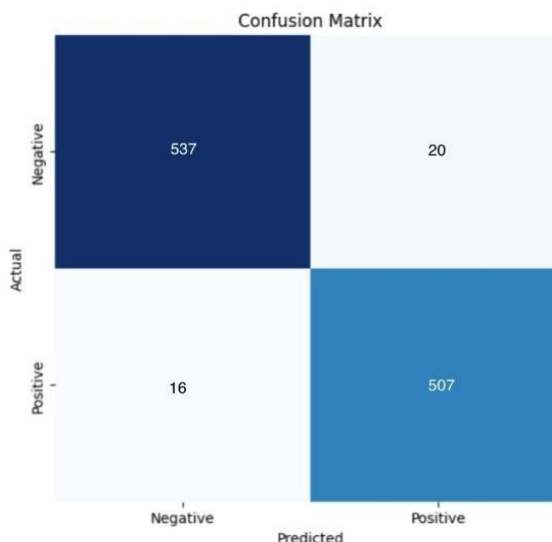
Evaluasi model merupakan bagian yang krusial dalam pengembangan model *deep learning* karena memungkinkan pengukuran performa model terhadap data yang belum pernah dilihat sebelumnya. Pada penelitian ini, beberapa metrik evaluasi digunakan untuk menilai model secara menyeluruh, termasuk *precision*, *recall*, *F1-score*, dan *accuracy*. Metrik-metrik tersebut memberikan gambaran yang jelas mengenai kemampuan model dalam melakukan klasifikasi sentimen ulasan produk pada platform *e-commerce* Tokopedia. Tabel 9 di bawah ini menyajikan hasil evaluasi model berdasarkan metrik yang telah disebutkan.

Tabel 9. Kinerja Model Klasifikasi

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
Negatif	97%	96%	97%	97%
Positif	96%	97%	97%	

Berdasarkan evaluasi yang disajikan pada Tabel 9, performa model menunjukkan hasil yang baik dalam mengklasifikasikan sentimen ulasan produk di platform Tokopedia. *Accuracy* model mencapai 97%, yang menggambarkan bahwa sebagian besar prediksi yang dihasilkan sesuai dengan label asli pada data uji. Dengan *precision* untuk kelas *negatif* sebesar 97%, model menunjukkan kemampuan yang baik dalam mengidentifikasi ulasan *negatif* dengan kesalahan yang minimal. Sementara itu, *recall* untuk kelas *negatif* sebesar 96% menunjukkan bahwa model mampu menangkap sebagian besar data yang memang termasuk dalam kelas *negatif*. Performa model pada kelas *positif* juga menunjukkan hasil yang seimbang, dengan *precision* sebesar 96% dan *recall* sebesar 97%. Ini menggambarkan bahwa model mampu mengidentifikasi ulasan *positif* dengan tingkat kesalahan yang rendah serta mencakup sebagian besar data yang relevan dengan kelas tersebut. Performa seimbang antara *precision* dan *recall* pada kedua kelas ini menunjukkan bahwa model tidak condong pada satu kelas tertentu, sehingga menghasilkan hasil klasifikasi yang proporsional dan akurat. Selain itu, nilai *F1-score* yang konsisten sebesar 97% pada kedua kelas memberikan indikasi bahwa model tidak hanya unggul dalam hal ketepatan (*precision*), tetapi juga sensitif dalam mengidentifikasi semua contoh yang termasuk dalam kelas tertentu (*recall*). Dengan demikian, *F1-score* yang stabil memperlihatkan bahwa model ini mampu menjaga keseimbangan antara kedua metrik tersebut, yang penting dalam konteks klasifikasi sentimen. Hasil evaluasi ini memberikan gambaran performa yang komprehensif dari model *IndoBERT* yang digunakan dalam klasifikasi sentimen ulasan produk di Tokopedia. *Accuracy* yang tinggi serta keseimbangan antara *precision*, *recall*, dan *F1-score* mengindikasikan bahwa model memiliki kemampuan yang baik dalam menangani variasi ulasan yang muncul di platform *e-commerce* Tokopedia.

Penelitian ini juga menggunakan *Confusion Matrix* sebagai metode untuk mengukur kinerja model klasifikasi. *Confusion Matrix* adalah metode yang sering digunakan dalam analisis performa model klasifikasi karena memberikan informasi yang rinci mengenai hasil prediksi model terhadap data uji. Matriks ini membagi hasil prediksi ke dalam empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dengan demikian, *Confusion Matrix* memungkinkan pengukuran kesalahan serta keberhasilan model dalam mengklasifikasikan setiap kelas. *True Positive* (TP) merujuk pada jumlah prediksi *positif* yang benar, yang berarti model berhasil memprediksi dengan tepat bahwa sampel termasuk dalam kelas *positif*. *True Negative* (TN) menunjukkan jumlah prediksi *negatif* yang benar, di mana model mampu mengidentifikasi sampel yang tidak termasuk dalam kelas *positif*. *False Positive* (FP) terjadi ketika model salah mengklasifikasikan sampel *negatif* sebagai *positif*, sementara *False Negative* (FN) adalah kasus di mana model salah memprediksi sampel *positif* sebagai *negatif*. Klasifikasi ini membantu dalam mengidentifikasi kekuatan dan kelemahan model dalam melakukan prediksi pada data uji.



Gambar 3. *Confusion Matrix*

Confusion Matrix pada Gambar 3 menunjukkan distribusi hasil prediksi model terhadap 1.080 sampel ulasan yang terbagi menjadi dua kategori: ulasan *positif* dan ulasan *negatif*. Berdasarkan evaluasi ini, dari 557 ulasan *negatif* yang ada, model berhasil memprediksi 537 ulasan dengan benar sebagai *negatif* (*True Negative*), sementara 20 ulasan *negatif* diprediksi salah sebagai *positif* (*False Positive*). Sebaliknya, dari 523 ulasan *positif*, 507 ulasan berhasil diprediksi dengan benar sebagai *positif* (*True Positive*), dan 16 ulasan salah diklasifikasikan sebagai *negatif* (*False Negative*). Hasil ini menunjukkan bahwa model mampu mengidentifikasi ulasan *negatif* dan *positif* dengan tingkat akurasi yang baik. Kesalahan prediksi, seperti yang terlihat dari jumlah *False Positive* dan *False Negative*, menunjukkan bahwa model masih memiliki beberapa kesalahan dalam mengklasifikasikan ulasan. Namun, kesalahan tersebut relatif kecil jika dibandingkan dengan jumlah total ulasan, yang mengindikasikan bahwa model ini bekerja dengan baik. Jumlah kesalahan prediksi yang relatif kecil memberikan dasar yang kuat untuk menyimpulkan bahwa model telah menunjukkan kinerja yang memadai dalam tugas klasifikasi ini.

3.4 Perbandingan Performa Model

Penelitian ini membandingkan kinerja model IndoBERT dengan model yang digunakan dalam penelitian sebelumnya yang menggunakan dataset *PRDECT-ID*, yaitu *LSTM* yang dikombinasikan dengan *embedding FastText*, *Glove*, dan *Word2Vec*. Penelitian sebelumnya menunjukkan bahwa kombinasi *LSTM* dengan *embedding FastText* mencapai akurasi tertinggi sebesar 85,08%, sementara kombinasi dengan *Word2Vec* dan *Glove* masing-masing mencapai akurasi sebesar 84,62% dan 83,04% [6]. Perbandingan ini dilakukan untuk menilai performa model IndoBERT dalam klasifikasi sentimen ulasan produk di Tokopedia dibandingkan pendekatan berbasis *LSTM* yang telah diterapkan sebelumnya. Hasil kinerja performa antara model *LSTM* dengan berbagai embedding dan model IndoBERT ditampilkan pada Tabel 10.

Tabel 10. Perbandingan Model

Model	Akurasi(%)
LSTM +FastText[6]	85,08%
LSTM + Word2Vec[6]	84,62%
LSTM + GloVe[6]	83,04%
IndoBERT	97%

Berdasarkan Tabel 10, terlihat bahwa model IndoBERT memberikan performa akurasi yang jauh lebih tinggi, mencapai 97%, dibandingkan dengan model *LSTM* yang menggunakan embedding *FastText*, *Word2Vec*, maupun *Glove* pada penelitian sebelumnya. Hasil ini mengkonfirmasi bahwa IndoBERT, sebagai model yang dirancang untuk memproses teks dalam bahasa Indonesia, mampu memberikan performa yang lebih unggul dalam memahami dan mengklasifikasikan sentimen ulasan produk di platform e-commerce seperti Tokopedia khususnya menggunakan dataset *PRDECT-ID*. IndoBERT, yang telah melalui proses pelatihan menggunakan korpus besar berbahasa Indonesia, menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen ulasan produk. Hal ini disebabkan oleh kemampuannya dalam menangkap konteks kalimat secara dinamis melalui arsitektur transformer dan representasi embedding kontekstual. Sebaliknya, model *LSTM* dengan embedding statis seperti *FastText*, *Word2Vec*, dan *Glove* memiliki keterbatasan dalam memahami hubungan antar kata yang lebih kompleks dalam teks. Dengan demikian, pemanfaatan *pre-training IndoBERT* pada korpus bahasa Indonesia dan penggunaan arsitektur transformer memungkinkan model ini untuk mengolah teks lebih efektif dibandingkan pendekatan *LSTM* dengan embedding statis, khususnya dalam tugas klasifikasi sentimen ulasan produk berbahasa Indonesia.

4. KESIMPULAN

Penelitian ini membandingkan performa model *IndoBERT* dengan pendekatan *LSTM* yang dikombinasikan dengan embedding *FastText*, *Word2Vec*, dan *GloVe* dalam klasifikasi sentimen ulasan produk di Tokopedia menggunakan dataset *PRDECT-ID*. Hasil penelitian menunjukkan bahwa *IndoBERT* memiliki akurasi lebih tinggi, yaitu 97% dalam memisahkan sentimen positif dan negatif, dibandingkan dengan model berbasis *LSTM*, yang memiliki akurasi tertinggi 85,08% dengan menggunakan embedding *FastText*. Performa *IndoBERT* yang lebih baik disebabkan oleh kemampuannya dalam menangkap konteks kata secara lebih mendalam melalui arsitektur transformer yang telah dilatih menggunakan korpus bahasa Indonesia. Sama seperti penelitian terdahulu, penelitian ini masih terbatas pada penggunaan dataset *PRDECT-ID*, sehingga hasil yang diperoleh hanya merepresentasikan karakteristik data tersebut. Selain itu, penelitian ini belum mengeksplorasi variasi dataset lain atau membandingkan *IndoBERT* dengan model *transformer* lain yang mungkin juga efektif dalam tugas klasifikasi sentimen. Untuk penelitian selanjutnya, disarankan untuk memperluas cakupan dataset serta mengeksplorasi teknik *fine-tuning* lebih lanjut untuk mengoptimalkan performa model dalam penelitian.

REFERENCES

- [1] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal Eksplora Informatika*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/EKSPLORA.V10I1.390.
- [2] R. Hidayansyah, A. Ahmad, and N. I. Nabila, "The impact of consumer reviews and ratings on purchase decisions on the Tokopedia marketplace in Indonesia," *International Journal of Economics, Business, and Entrepreneurship*, vol. 6, no. 2, pp. 140–152, Dec. 2023, doi: 10.23960/IJEBE.V6I2.264.
- [3] A. N. Rohman, R. Luviana Musyarofah, E. Utami, and S. Raharjo, "Natural Language Processing on Marketplace Product Review Sentiment Analysis," in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2020, pp. 1–5. doi: 10.1109/ICORIS50180.2020.9320827.
- [4] R. Hermansyah and R. Sarno, "Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method," in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 511–516. doi: 10.1109/iSemantic50169.2020.9234238.
- [5] D. Sharma and M. Sabharwal, "Sentiment analysis for social media using SVM classifier of machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9 Special Issue 4, pp. 39–47, Jul. 2019, doi: 10.35940/IJITEE.I1107.0789S419.
- [6] N. Dirfas and V. Nastiti, "Perbandingan Kinerja Pre-Trained Word Embedding Terhadap Performa Klasifikasi Sentimen Ulasan Produk Tokopedia Dengan Long Short-Term Memory(LSTM)," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, p. 878–889, Sep. 2024, doi: 10.47065/bits.v6i2.5634.
- [7] M. V. Koroteev, "BERT: a review of applications in natural language processing and understanding," *arXiv preprint arXiv:2103.11943*, 2021.
- [8] M. P. Geetha and D. Karthika Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model," *International Journal of Intelligent Networks*, vol. 2, pp. 64–69, Jan. 2021, doi: 10.1016/J.IJIN.2021.06.005.
- [9] N. N. Qomariyah, T. Sun, and D. Kazakov, "NLP Analysis of COVID-19 Radiology Reports in Indonesian using IndoBERT," in *2022 4th International Conference on Biomedical Engineering (IBIOMED)*, 2022, pp. 65–70. doi: 10.1109/IBIOMED56408.2022.9988223.
- [10] R. Sutoyo, S. Achmad, A. Chowanda, E. W. Andangsari, and S. M. Isa, "PRDECT-ID: Indonesian product reviews dataset for emotions classification tasks," *Data Brief*, vol. 44, p. 108554, Oct. 2022, doi: 10.1016/J.DIB.2022.108554.
- [11] J. Liu *et al.*, "Application of Deep Learning-Based Natural Language Processing in Multilingual Sentiment Analysis," *Mediterranean Journal of Basic and Applied Sciences*, vol. 08, no. 02, pp. 243–260, 2024, doi: 10.46382/MJBAS.2024.8219.
- [12] C. C. P. Hapsari, W. Astuti, and M. D. Purbolaksono, "Naive Bayes Classifier and Word2Vec for Sentiment Analysis on Bahasa Indonesia Cosmetic Product Reviews," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*, 2021, pp. 22–27. doi: 10.1109/ICoDSA53588.2021.9617544.
- [13] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 2020, pp. 217–220. doi: 10.1109/IC3A48958.2020.233300.
- [14] L. Mathew and V. R. Bindu, "A Review of Natural Language Processing Techniques for Sentiment Analysis using Pre-trained Models," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 2020, pp. 340–345. doi: 10.1109/ICCMC48092.2020.ICCMC-00064.
- [15] Y. Xu and R. Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," *J Anal Test*, vol. 2, no. 3, pp. 249–262, Jul. 2018, doi: 10.1007/S41664-018-0068-2.
- [16] S. Ravichandiran, *Getting Started with Google BERT: Build and train state-of-the-art natural ... - Sudharsan Ravichandiran - Google Books*. 2021. Accessed: Feb. 20, 2025. [Online]. Available: https://books.google.co.id/books?hl=en&lr=&id=CvsWEAAAQBAJ&oi=fnd&pg=PP1&dq=Getting+started+with+Google+BERT+:+build+and+train+state-of-the-art+natural+language+processing+models+using+BERT&ots=3HiCs4le62&sig=2KnOmeWWcGhPLwa437c9JjOAWQ&redir_esc=y#v=onepage&q=Getting%20started%20with%20Google%20BERT%20%3A%20build%20and%20train%20state-of-the-art%20natural%20language%20processing%20models%20using%20BERT&f=false



- [17] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2018, Accessed: Feb. 20, 2025. [Online]. Available: <https://arxiv.org/abs/1810.04805v2>
- [18] A. Nayak, H. Timmapathini, K. Ponnalagu, and V. Gopalan Venkoparao, “Domain adaptation challenges of BERT in tokenization and sub-word representations of Out-of-Vocabulary words,” in *Proceedings of the First Workshop on Insights from Negative Results in NLP*, A. Rogers, J. Sedoc, and A. Rumshisky, Eds., Nov. 2020, pp. 1–5. doi: 10.18653/v1/2020.insights-1.1.
- [19] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 757–770, Nov. 2020, doi: 10.18653/v1/2020.coling-main.66.
- [20] H. D. Sharma and P. Goyal, “An Analysis of Sentiment: Methods, Applications, and Challenges,” *Engineering Proceedings*, vol. 59, no. 1, 2023, doi: 10.3390/engproc2023059068.
- [21] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken DW, F. A. Bachtiar, and N. Yudistira, “BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews,” in *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology*, in SIET '21. Association for Computing Machinery, 2021, pp. 258–264. doi: 10.1145/3479645.3479679.
- [22] L. R. Aini, E. Nurfadhilah, A. Jarin, A. Santosa, and M. T. Uliniansyah, “Enhancing Sentiment Analysis Models through Multi-Technique Data Augmentation: A Study with IndoBERT,” in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 2023, pp. 137–142. doi: 10.1109/IC3INA60834.2023.10285775.
- [23] B. Agarwal, R. Nayak, N. Mittal, and S. Patnaik, Eds., *Deep Learning-Based Approaches for Sentiment Analysis. in Algorithms for Intelligent System* s. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-15-1216-2.
- [24] S. Saadah, K. M. Auditama, A. A. Fattahila, F. I. Amorokhman, A. Aditsania, and A. A. Rohmawati, “Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, Aug. 2022, doi: 10.29207/RESTI.V6I4.4215.
- [25] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, “Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning,” *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.