Journal of Digital Business and Technology Innovation (DBESTI)

https://journal.nurulfikri.ac.id/index.php/DBESTI
P-ISSN: 3047-1028 E-ISSN: 3032-775X

ANALISIS SENTIMEN ULASAN PRODUK ELEKTRONIK TOKOPEDIA MENGGUNAKAN *BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS* (BERT)

Aliyah Qulbinisa Setiyawan[1], Ahmad Rio Adriansyah[2], Tubagus Rizky Dharmawan[3]

[1,2,3]Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri
Jakarta Selatan, DKI Jakarta, Indonesia 12640
aliy21019ti@student.nurulfikri.ac.id, arasy@nurulfikri.ac.id, tubagus@nurulfikri.ac.id

Abstract

The growth of e-commerce in Indonesia, especially Tokopedia, has led to many consumer reviews containing important opinions, particularly in the electronics category. Manually analyzing these reviews is inefficient, so an automated approach is needed. This study uses the Bidirectional Encoder Representations from Transformers (BERT) model, specifically the IndoBERT variant, to analyze sentiment in customer reviews of electronics products on Tokopedia. The research stages include preprocessing (cleaning, tokenizing, stopword removal, stemming), sentiment labeling (positive, negative, neutral), model training, and evaluation with accuracy, precision, recall, and F1-score metrics. The results show that the model performs very well, with an accuracy of 99.69% and precision and recall values of 1.00 for positive and neutral sentiments. Negative sentiment has a recall of 0.94, influenced by the imbalanced data distribution. The main factors that influence user sentiment include product quality, delivery speed, and product description match. This method is effective for sentiment classification in Indonesian.

Keywords: *BERT, Electronic Products, IndoBERT, Sentiment Analysis, Tokopedia*

Abstrak

Pertumbuhan *e-commerce* di Indonesia, khususnya Tokopedia, telah memunculkan banyak ulasan konsumen yang berisi opini-opini penting, terutama dalam kategori produk elektronik. Melakukan analisis secara manual terhadap ulasan-ulasan tersebut dinilai kurang efisien, sehingga diperlukan pendekatan otomatis. Penelitian ini memanfaatkan model *Bidirectional Encoder Representations from Transformers* (BERT), khususnya varian IndoBERT, untuk melakukan analisis sentimen terhadap ulasan pelanggan pada kategori produk elektronik di Tokopedia. Tahapan penelitian mencakup *preprocessing* (*cleaning*, *tokenizing*, *stopword removal*, *stemming*), pelabelan sentimen (positif, negatif, netral), pelatihan model, dan evaluasi dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan performa model sangat baik, dengan akurasi 99,69% dan nilai presisi serta *recall* sebesar 1.00 untuk sentimen positif dan netral. Sentimen negatif memiliki *recall* 0.94 dipengaruhi oleh ketidakseimbangan distribusi data. Faktor utama yang memengaruhi sentimen pengguna meliputi kualitas produk, kecepatan pengiriman, dan kecocokan deskripsi produk. Metode ini efektif untuk klasifikasi sentimen dalam bahasa Indonesia.

Kata kunci: Analisis Sentimen, BERT, IndoBERT, Produk Elektronik, Tokopedia

1. PENDAHULUAN

Dengan berkembangnya teknologi digital, khususnya platform *e-commerce*, telah terjadi perubahan besar dalam aktivitas perdagangan. Tokopedia menjadi salah satu platform perdagangan *online* terbesar di Indonesia, melayani jutaan transaksi produk setiap hari, termasuk dalam kategori produk elektronik. Setiap transaksi biasanya memiliki ulasan atau komentar pengguna, yang berisi informasi penting tentang pengalaman pengguna dengan barang dan layanan yang diterima.

Untuk memahami persepsi dan kepuasan pelanggan, ulasan pelanggan menjadi sumber data yang sangat berharga bagi bisnis dan penyedia platform. Namun, karena volume dan keanekaragaman ulasan yang sangat besar, proses analisis pelanggan secara manual menjadi tidak efisien dan rentan terhadap bias. Oleh karena itu, dibutuhkan metode analisis otomatis yang mampu mengolah data teks dalam jumlah besar dan memahami makna yang terkandung dalam ulasan tersebut secara kontekstual.

Diterima 14 Juli 2025, Direvisi 24 November 2026, Diterima untuk publikasi 24 November 2026

Bidirectional Encoder Representations from Transformers (BERT) merupakan salah satu pendekatan berbasis *deep learning* yang telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami (*Natural Language Processing*), termasuk klasifikasi sentimen. Keunggulan utama dari BERT adalah kemampuannya dalam memahami konteks kata dari dua arah secara simultan, sehingga sangat cocok untuk menganalisis opini atau sentimen dalam teks[1]. Di Indonesia, pengembangan model BERT yang diadaptasi untuk Bahasa Indonesia dikenal dengan IndoBERT, yang telah dilatih pada korpus bahasa lokal dan menunjukkan performa baik dalam tugas klasifikasi teks[2].

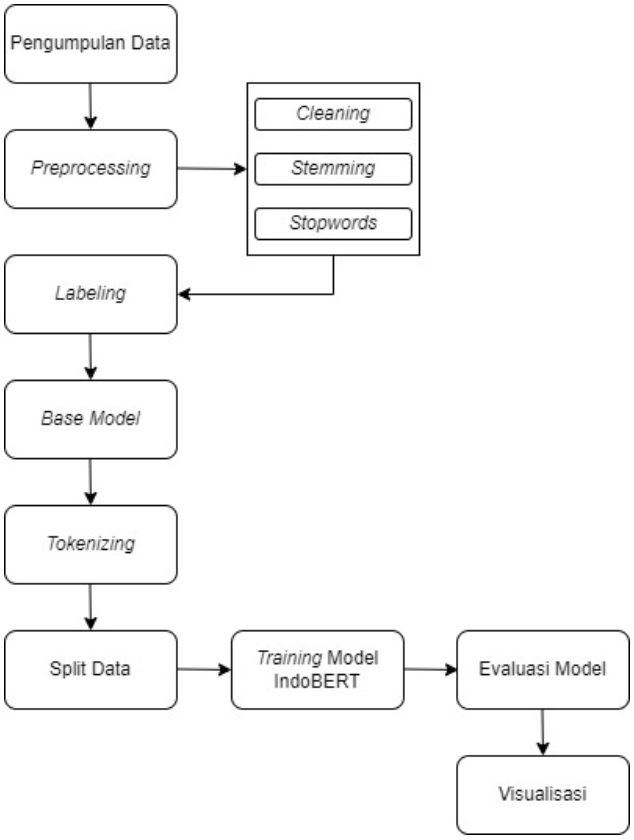
Penelitian ini bertujuan untuk menerapkan model IndoBERT dalam menganalisis sentimen pada ulasan pelanggan terhadap produk elektronik di Tokopedia. Produk elektronik dipilih karena merupakan salah satu kategori dengan volume transaksi dan ulasan yang tinggi, serta memiliki variabilitas opini yang luas, mulai dari aspek teknis hingga layanan pengiriman. Dengan mengklasifikasikan ulasan menjadi sentimen positif, negatif, dan netral, penelitian ini diharapkan dapat memberikan kontribusi dalam mendukung pengambilan keputusan berbasis data, baik bagi penyedia platform, produsen, maupun calon konsumen.

Penerapan model IndoBERT dalam konteks ini tidak hanya bertujuan untuk mengukur efektivitas teknologi NLP dalam bahasa Indonesia, tetapi juga untuk mengidentifikasi pola-pola utama dalam persepsi pelanggan terhadap produk elektronik. Hasil penelitian ini diharapkan dapat menjadi rujukan dalam pengembangan sistem rekomendasi produk, peningkatan layanan pelanggan, serta strategi pemasaran berbasis analisis opini secara otomatis dan cerdas.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan memanfaatkan teknik *Natural Language Processing* (NLP) yang terintegrasi dalam model BERT (*Bidirectional Encoder Representations from Transformers*). Tujuan dari pendekatan ini adalah untuk menggambarkan pola sentimen pelanggan dalam bentuk data numerik yang diperoleh melalui hasil klasifikasi model terhadap ulasan pelanggan pada platform *e-commerce* Tokopedia. Sentimen yang diklasifikasikan mencakup tiga kategori, yaitu positif, negatif, dan netral, berdasarkan opini pengguna terhadap produk elektronik.

Pendekatan kuantitatif dalam penelitian ini memungkinkan pengukuran frekuensi, distribusi, dan intensitas dari setiap kategori sentimen yang dianalisis. Hasil klasifikasi sentimen ini memberikan gambaran numerik mengenai persepsi konsumen terhadap kualitas produk, kecepatan pengiriman, kesesuaian deskripsi, dan aspek pelayanan lainnya. Selain itu, metode ini juga memungkinkan evaluasi performa model secara objektif menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*, sehingga validitas hasil dapat dipertanggung jawabkan secara ilmiah[3].



Gambar 1. Tahapan Penelitian

Gambar 1 menjelaskan alur proses kerja sistem analisis sentimen menggunakan model IndoBERT yang diimplementasikan dalam penelitian ini. Proses dimulai dengan pengumpulan data berupa ulasan pelanggan terhadap produk elektronik di platform Kaggle. Data tersebut dikumpulkan sebagai masukan utama yang akan diproses lebih lanjut. Setelah data terkumpul, tahap selanjutnya adalah *preprocessing*, yang mencakup tiga proses inti, yaitu *cleaning*, *stemming*, *stopwords*.

Setelah proses pembersihan teks selesai, data dilabeli ke dalam tiga kategori sentimen: positif, negatif, dan netral. Pelabelan ini bertujuan untuk memberikan target kelas bagi model klasifikasi. Data yang telah diberi label kemudian masuk ke tahap *base model*, di mana IndoBERT disiapkan sebagai model dasar yang sudah dilatih sebelumnya (*pre-trained*). Tahap berikutnya adalah tokenisasi, yaitu mengubah teks ke dalam bentuk token ID numerik menggunakan *tokenizer* khusus dari IndoBERT, agar bisa dikenali dan diproses oleh arsitektur *transformer*.

Setelah data ditokenisasi, dilakukan pembagian data (*split data*) menjadi dua subset, yaitu data latih (*training* data) dan data uji (*testing* data). Proses pelatihan dilakukan menggunakan model IndoBERT terhadap data latih dalam beberapa *epoch*, dengan konfigurasi *batch size* dan *learning rate* tertentu. Hasil pelatihan model kemudian dievaluasi menggunakan data uji, dengan mengukur performa model melalui metrik klasifikasi seperti akurasi, presisi, *recall*, dan *F1-score*. Terakhir, hasil evaluasi divisualisasikan dalam

bentuk grafik distribusi sentimen, *confusion matrix*, dan *word cloud*, guna memberikan interpretasi yang lebih mudah dipahami terhadap kinerja model dan persebaran opini pelanggan.

2.1 Pengumpulan Data

Data dalam penelitian ini diperoleh melalui metode studi literatur dan pemanfaatan data sekunder dari sumber terbuka. Dataset yang digunakan berasal dari situs Kaggle, dengan judul **"**Tokopedia *Product Reviews***"**, yang berisi kumpulan ulasan pelanggan terhadap berbagai produk, khususnya dalam kategori elektronik. Dataset ini mencakup beberapa atribut penting, seperti teks ulasan (*review text*), peringkat bintang (*rating*), nama produk, serta kategori produk. Pemanfaatan data sekunder dari sumber pihak ketiga seperti Kaggle memungkinkan peneliti untuk memperoleh data dalam skala besar dan terstruktur, yang sangat mendukung penerapan teknik analisis berbasis *machine learning* dan *deep learning*.

2.2 *Preprocessing*

Preprocessing merupakan proses penting yang bertujuan untuk mengubah data mentah menjadi bentuk yang lebih terstruktur dan mudah diolah oleh sistem. Tahapan ini dilakukan untuk mengatasi berbagai permasalahan umum dalam data teks, seperti ketidakakuratan, *redundansi*, serta keberadaan data yang tidak lengkap atau tidak relevan. Dengan menerapkan *preprocessing* secara tepat, kualitas data dapat ditingkatkan, sehingga model klasifikasi dapat bekerja secara lebih efisien dan fokus pada informasi yang benar-benar penting, tanpa terganggu oleh gangguan (*noise*) dari elemen-elemen data yang tidak diperlukan. Ada tiga tahapan *preprocessing* yang digunakan, sebagai berikut:

a) *Stemming*

Proses penghilangan afiks atau imbuhan dari suatu kata dengan tujuan untuk mengembalikan kata tersebut ke bentuk dasarnya. Proses ini bertujuan untuk membantu model dalam mengelompokkan berbagai bentuk turunan dari satu kata dasar sebagai representasi yang sama. Dengan demikian, model dapat melakukan generalisasi yang lebih baik terhadap data pelatihan. Sebagai contoh, kata "berjalan" akan diproses menjadi bentuk dasarnya, yaitu "jalan"[4].

b) *Stopword*

Teknik yang digunakan untuk menghapus kata-kata umum yang sering muncul dalam bahasa, namun tidak memiliki makna penting dalam konteks analisis. Kata-kata seperti "dan", "di", "ke", atau "yang" biasanya tidak memberikan kontribusi signifikan terhadap pemahaman isi teks. Dengan menghilangkan *stopwords*, proses analisis dapat difokuskan pada kata-kata utama yang memiliki bobot informasi lebih besar dalam menentukan konteks atau sentimen dari suatu kalimat[4].

c) *Tokenizing*

Proses membagi teks menjadi unit-unit kecil yang disebut token, yang dapat berupa kata, frasa, atau simbol. Langkah ini mempermudah proses analisis karena memungkinkan sistem untuk mengidentifikasi dan menghitung frekuensi kemunculan kata-kata dalam dokumen. Dengan tokenisasi, model dapat memetakan struktur teks dan mengenali pola penggunaan bahasa yang muncul dalam kumpulan data[4].

2.3 *Labeling*

Setelah data teks selesai diproses melalui tahap *preprocessing*, langkah berikutnya adalah *labeling*, yaitu pemberian label sentimen (positif, negatif, atau netral) pada setiap ulasan. Dalam penelitian ini, *labeling* dilakukan dengan menggunakan metode *lexicon-based* sederhana, di mana setiap kata dalam teks dicocokkan dengan kamus sentimen yang berisi daftar kata positif dan negatif.

Metode *lexicon-based labeling* telah banyak digunakan dalam penelitian analisis sentimen Bahasa Indonesia. Sebagai contoh, sebuah studi *review* menyebutkan bahwa proses klasifikasi didasarkan pada perbandingan jumlah kata sentimen negatif dan positif dalam sebuah kalimat, serta menunjukkan bahwa metode ini mampu menghasilkan akurasi hingga 90% pada berbagai dataset Indonesia[5].

2.4 *Base Model*

Menentukan model dasar (*base model*) yang digunakan dalam proses klasifikasi. Penelitian ini menggunakan IndoBERT, yaitu hasil *pre-training* dari BERT (*Bidirectional Encoder Representations from Transformers*) yang disesuaikan untuk bahasa Indonesia. IndoBERT dipilih karena telah dilatih pada korpus besar berbahasa Indonesia, sehingga memiliki pemahaman semantik dan sintaksis yang lebih baik untuk konteks lokal[1].

Model IndoBERT dapat langsung digunakan sebagai *feature extractor* atau di-*fine-tune* kembali agar sesuai dengan domain dataset ulasan Tokopedia. Penggunaan model *pre-trained* ini mempercepat proses pengembangan dan meningkatkan akurasi klasifikasi karena sudah memahami struktur bahasa Indonesia dengan baik[6].

2.5 *Tokenizing*

Setelah model dimuat, tahapan selanjutnya adalah tokenisasi data teks ulasan. Tokenisasi dilakukan dengan *tokenizer* khusus IndoBERT (IndoBERT *tokenizer*) yang mengonversi kalimat teks menjadi format vektor numerik (input_ids, attention_mask, dan token_type_ids) yang dapat diproses oleh model BERT[7].

Selain itu, label sentimen yang semula berupa teks seperti "positif", "negatif", dan "netral", dikonversi menggunakan *LabelEncoder* ke dalam bentuk numerik agar dapat dikenali oleh fungsi *loss* selama pelatihan.

2.6 *Split Data*

Data yang telah ditokenisasi kemudian dibagi menjadi dua bagian, yaitu data latih dan data validasi. Pembagian dilakukan dengan proporsi 80:20 menggunakan metode *stratified splitting*, agar distribusi label pada data pelatihan dan validasi tetap seimbang[8].

Pembagian data yang seimbang penting untuk mencegah bias model terhadap kelas tertentu, terutama karena distribusi data pada dataset ulasan Tokopedia cenderung tidak seimbang, di mana ulasan netral lebih dominan.

2.7 *Training*

Model dilatih menggunakan *framework* PyTorch dengan dataset hasil tokenisasi yang dikemas dalam objek Dataset dan DataLoader. Pelatihan dilakukan selama beberapa *epoch* (4–5) dengan *batch size* sebesar 16 dan *optimizer* AdamW. Untuk menghindari *overfitting*, diterapkan teknik *early stopping* dengan memantau nilai *validation loss* pada setiap *epoch*. Jika tidak ada peningkatan performa selama beberapa *epoch* berturut-turut, pelatihan dihentikan secara otomatis.

2.8 Evaluasi Model

Setelah data diklasifikasikan, metode selanjutnya adalah evaluasi. Metode evaluasi sangat penting untuk memastikan bahwa model dapat berfungsi dengan baik dengan data yang diujikan. Metode ini menerapkan model *Precision, Recall, Accuracy, F1-Score*.

a) *Precision*

Precision digunakan untuk mengukur ketepatan model dalam menentukan kategori. Dalam konteks ini, *True Positive* (TP) menunjukkan jumlah ulasan positif yang berhasil diklasifikasikan dengan tepat, sedangkan *False Positive* (FP) menunjukkan jumlah ulasan negatif yang keliru diprediksi sebagai positif. Semakin tinggi nilai ketepatan, semakin akurat model dalam memberikan label positif pada ulasan yang sebenarnya benar-benar positif. Rumus berikut dapat digunakan untuk menghitung *precision*[9]:

$$P = \frac{TP}{TP+FP}$$

b) *Recall*

Recall menilai kemampuan model untuk menemukan semua ulasan yang termasuk dalam kategori positif. Dalam hal ini, *True Positive* (TP) menunjukkan ulasan positif yang berhasil dikenali oleh model, sementara *False Negative* (FN) menunjukkan ulasan positif yang tidak berhasil ditemukan dan diklasifikasikan dengan benar. Nilai *recall* menunjukkan seberapa baik model menangkap semua data yang seharusnya termasuk dalam kelas positif. *Recall* dapat dihitung dengan rumus berikut[9]:

$$R = \frac{TP}{TP+FN}$$

c) *Accuracy*

Akurasi menunjukkan persentase prediksi yang tepat dibandingkan dengan total jumlah data yang diuji; dengan kata lain, akurasi digunakan untuk menilai seberapa besar proporsi prediksi yang dilakukan model sesuai dengan label sebenarnya dari seluruh data yang tersedia. Rumus berikut digunakan untuk menghitung nilai ini[6]:

$$A = \frac{TP+TN}{TP+TN+FP+FN}$$

d) *F1-Score*

F1-Score adalah metrik yang memberikan gambaran menyeluruh tentang kinerja model dengan menggabungkan nilai *recall* dan *precision* ke dalam satu nilai total. Metrik ini sangat berguna ketika ada ketidakseimbangan antara nilai *recall* dan *precision* karena mampu mengimbangi keduanya. Rumus berikut digunakan untuk melakukan perhitungan[9]:

$$F1 = 2\frac{r*p}{r+p}$$

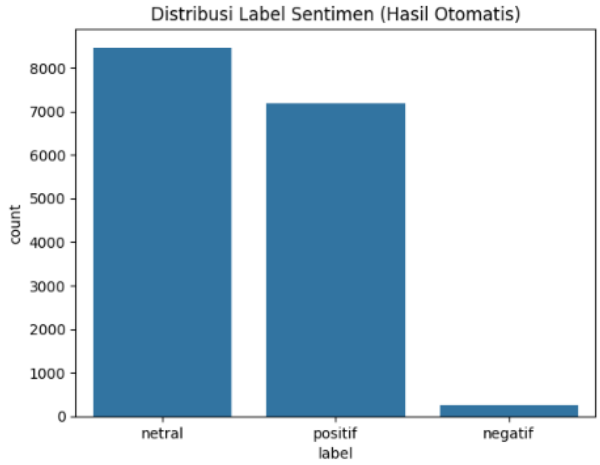
2.9 Visualisasi

Untuk memperjelas hasil analisis sentimen, dilakukan visualisasi berupa *word cloud* untuk menampilkan kata-kata dominan pada tiap sentimen, *confusion matrix* untuk melihat akurasi klasifikasi per kelas, serta grafik *training* dan *validation loss* guna memantau kestabilan pelatihan. Visualisasi ini dibuat menggunakan pustaka Matplotlib dan Seaborn.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari implementasi dan evaluasi model IndoBERT dalam menganalisis sentimen pelanggan terhadap produk elektronik di Tokopedia. Seluruh tahapan yang telah dijelaskan sebelumnya, mulai dari *preprocessing*, pelabelan data, pemilihan model, pelatihan, hingga evaluasi kinerja model, diuraikan secara sistematis untuk menunjukkan efektivitas pendekatan yang digunakan. Pembahasan difokuskan pada akurasi model dalam mengklasifikasikan ulasan ke dalam kategori sentimen positif, negatif, dan netral, serta bagaimana model merespons terhadap ketidakseimbangan data. Selain itu, visualisasi data ditampilkan untuk memperkuat interpretasi hasil dan menggambarkan persebaran kata kunci yang mencerminkan sentimen pelanggan.

Gambar 2 menunjukkan hasil distribusi label sentimen yang diperoleh melalui proses pelabelan otomatis berbasis leksikon terhadap ulasan pelanggan produk elektronik di Tokopedia. Ulasan diklasifikasikan ke dalam tiga kategori sentimen, yaitu netral, positif, dan negatif.




```
Distribusi label sentimen:
label
netral     8461
positif    7169
negatif     267
```

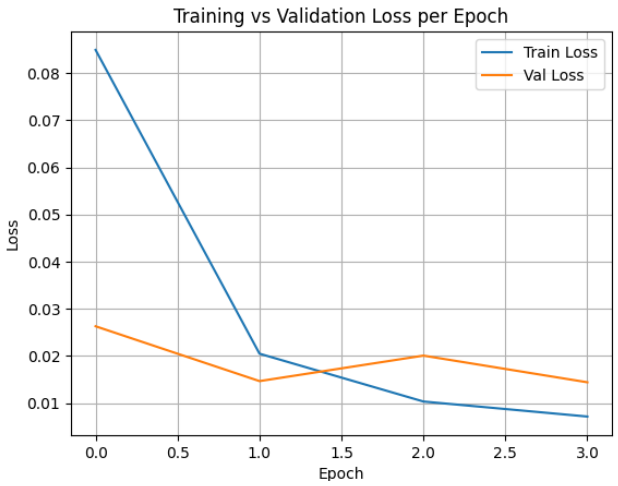
Gambar 2. Distribusi *Labeling*

Berdasarkan grafik, diketahui bahwa ulasan dengan sentimen netral mendominasi dengan jumlah sebanyak 8.461 ulasan, diikuti oleh ulasan positif sebanyak 7.169, sedangkan ulasan negatif hanya berjumlah 267. Ketimpangan distribusi ini mengindikasikan bahwa sebagian besar pelanggan menyampaikan ulasan yang bersifat informatif tanpa ekspresi emosional yang kuat, sedangkan ulasan negatif jumlahnya sangat sedikit.

```
1. Teks: Barang sesuai gambar, pengiriman sesuai jadwal...
   Prediksi: netral | Probabilitas: [[0.25309616327285767, 0.39030778408050537, 0.35659605264663696]]
2. Teks: barang cepat sampai, walau L2 agak keras overall good.. semoga awet!...
   Prediksi: netral | Probabilitas: [[0.18855243921279907, 0.4324389398097992, 0.3790086507797241]]
3. Teks: Produk dan pengiriman baik...
   Prediksi: positif | Probabilitas: [[0.23688599467277527, 0.34363245964050293, 0.4194815754890442]]
4. Teks: barang sesuai, bekerja dengan baik...
   Prediksi: netral | Probabilitas: [[0.2700258493423462, 0.4069114029407501, 0.32306280732154846]]
5. Teks: bingung install softwarenya gak komplit atau memang gak support di windows 8...
   Prediksi: netral | Probabilitas: [[0.28510016202926636, 0.37472978234291077, 0.3401700556278229]]
```

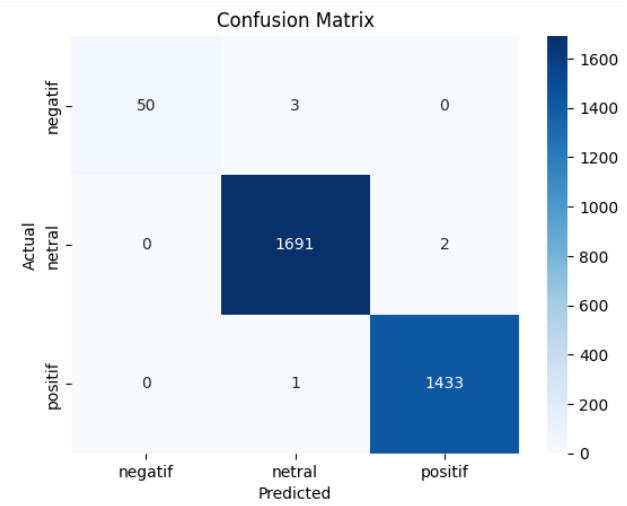
Gambar 3. *Base Model*

Hasil prediksi *base model* IndoBERT terhadap ulasan pelanggan ditunjukkan pada Gambar 3. Sentimen, seperti prediksi netral yang informatif dan positif dengan pernyataan kepuasan, dapat diidentifikasi dengan sukses pada setiap teks. Meskipun model dasar (*pre-trained*) IndoBERT menunjukkan kemampuan awal untuk memahami konteks kalimat dengan baik, hasil prediksi masih tidak konsisten. Ini terutama berlaku untuk ulasan yang mengandung makna campuran atau implisit. Hal ini menunjukkan bahwa model cenderung memberikan prediksi yang kurang akurat jika tidak dilatih dengan proses pelatihan tambahan (*fine-tuning*) pada data domain tertentu.



Gambar 4. Grafik Perbandingan *Training Loss* dan *Validation Loss*

Gambar 4 menunjukkan grafik perbandingan antara *training loss* dan *validation loss* sepanjang proses pelatihan selama tiga *epoch*. Pada awal pelatihan (*epoch* 0), nilai *training loss* sangat tinggi, namun mengalami penurunan tajam pada *epoch* berikutnya. *Validation loss* juga menunjukkan penurunan dan tetap stabil di bawah 0.02, menandakan bahwa model tidak mengalami *overfitting*. Penurunan yang konsisten pada grafik tersebut menunjukkan bahwa parameter pelatihan seperti *batch size*, *learning rate*, dan jumlah *epoch* telah dikonfigurasi dengan tepat.



Gambar 5. *Heatmap*

Gambar 5 memperlihatkan *confusion matrix* yang menggambarkan hasil klasifikasi sentimen oleh model IndoBERT terhadap 3.180 data uji. Matriks ini menunjukkan bahwa model berhasil mengklasifikasikan 50 dari 53 data negatif dengan benar, sementara 3 sisanya diklasifikasikan sebagai netral. Untuk sentimen netral, model mencatat 1.691 prediksi benar dari 1.693 data, dengan hanya 2 data yang salah klasifikasi. Sementara untuk sentimen positif, 1.433 dari 1.434 data berhasil diklasifikasikan dengan tepat. Secara keseluruhan, *confusion matrix* menunjukkan bahwa model memiliki presisi tinggi dalam membedakan sentimen, dengan

kesalahan klasifikasi yang sangat kecil dan tidak signifikan terhadap performa keseluruhan.

```
Evaluasi Model Terbaik:

Classification Report:
              precision    recall  f1-score   support

     negatif       1.00      0.94      0.97        53
      netral       1.00      1.00      1.00      1693
     positif       1.00      1.00      1.00      1434

    accuracy                           1.00      3180
   macro avg       1.00      0.98      0.99      3180
weighted avg       1.00      1.00      1.00      3180

Akurasi Total: 0.9969
```

Gambar 6. Evaluasi Model

Evaluasi performa model ditunjukkan pada Gambar 6 dalam bentuk *classification report* yang memuat metrik *precision*, *recall*, *f1-score*, dan *accuracy*. Model IndoBERT menunjukkan kinerja yang sangat tinggi dengan akurasi total sebesar 99,69%. Untuk kategori sentimen positif dan netral, nilai *precision*, *recall*, dan *f1-score* mencapai 1.00, menandakan bahwa model sangat akurat dalam memprediksi dua kelas tersebut. Namun, untuk kategori negatif, *recall* berada pada angka 0.94, sedikit lebih rendah dibandingkan kelas lainnya akibat jumlah data yang jauh lebih sedikit. Nilai *macro average f1-score* sebesar 0.99 mengindikasikan bahwa performa antar kelas relatif seimbang meskipun ada ketimpangan data.

Untuk mendapatkan pemahaman yang lebih baik tentang konteks dan konten dari setiap kategori sentimen, dilakukan visualisasi *cloud* kata terhadap ulasan pelanggan yang dikategorikan ke dalam tiga kategori: positif, negatif, dan netral. Tujuan visualisasi ini adalah untuk menemukan kata-kata yang paling sering muncul terkait dengan masing-masing kategori sentimen, sehingga dapat memberikan gambaran luas tentang topik utama yang disampaikan pelanggan[10].



Gambar 7. Kata-kata yang terdapat pada ulasan positif

Gambar 7 menunjukkan *word cloud* untuk sentimen positif. Kata-kata yang paling dominan adalah “cepat”, “barang”, “bagus”, “sesuai”, dan “respon”. Ini menandakan bahwa kepuasan pelanggan banyak berkaitan dengan kecepatan pengiriman, kualitas produk, serta kesesuaian produk dengan deskripsi. Selain itu, munculnya kata seperti “mantap”, “terima kasih”, dan “recommended” semakin menguatkan kesan positif yang ditunjukkan oleh pelanggan terhadap pengalaman belanja mereka. Kata-kata ini menggambarkan persepsi yang baik terhadap layanan dan produk yang diterima.



Gambar 8. Kata-kata yang terdapat pada ulasan negatif

Gambar 8 menampilkan *word cloud* untuk sentimen negatif. Kata-kata seperti “rusak”, “gak”, “tidak sesuai”, dan “mengecewakan” menjadi indikator kuat adanya ketidakpuasan pelanggan. Meskipun beberapa kata seperti “terima kasih” dan “barang” juga muncul, namun konteks penggunaannya cenderung bersifat sarkastik atau beriringan dengan keluhan. Kata “alamat”, “salah”, dan “fungsi” juga muncul, yang mengindikasikan keluhan seputar kerusakan produk, kesalahan pengiriman, atau barang yang tidak berfungsi dengan baik.



Gambar 9. Kata-kata yang terdapat pada ulasan netral

Gambar 9 menampilkan *word cloud* dari ulasan dengan sentimen netral. Kata-kata seperti “barang”, “sesuai”, “terima kasih”, “deskripsi”, dan “fungsi” mendominasi, namun cenderung tidak menyampaikan emosi yang kuat baik positif maupun negatif. Kata-kata tersebut lebih bersifat informatif, menggambarkan kondisi produk secara objektif tanpa opini yang menilai. Hal ini mengindikasikan bahwa pelanggan dalam kategori ini lebih fokus pada pelaporan pengalaman atau status pesanan tanpa memberikan ekspresi kepuasan atau ketidakpuasan secara eksplisit.

4. KESIMPULAN

Hasil analisis sentimen menggunakan model IndoBERT menunjukkan bahwa mayoritas ulasan pelanggan

mengandung sentimen netral dan positif, sedangkan sentimen negatif jumlahnya relatif sedikit. Hal ini didukung oleh hasil evaluasi dan *confusion matrix* yang menunjukkan dominasi prediksi pada kedua kelas tersebut. Visualisasi *word cloud* juga memperkuat temuan ini, di mana kata-kata seperti "bagus", "cepat", "berfungsi", dan "sesuai" mendominasi sentimen positif, mencerminkan kepuasan terhadap kualitas produk dan layanan pengiriman. Sebaliknya, kata-kata seperti “rusak”, “lambat”, dan “mengecewakan” muncul pada sentimen negatif, menunjukkan ketidakpuasan terhadap produk dan layanan. Faktor utama pembentuk sentimen secara umum meliputi kualitas produk, kecepatan pengiriman, dan kesesuaian deskripsi dengan ekspektasi. Secara keseluruhan, IndoBERT terbukti efektif dalam memahami konteks bahasa Indonesia dan menghasilkan klasifikasi sentimen yang akurat. Model ini mencapai akurasi sebesar 99,69% dengan nilai *precision*, *recall*, dan *F1-score* yang tinggi, terutama pada kelas positif dan netral. Sementara itu4r, *recall* untuk sentimen negatif sedikit lebih rendah, yakni 0,94, yang disebabkan oleh jumlah data negatif yang jauh lebih sedikit dibandingkan dua kelas lainnya.

DAFTAR PUSTAKA

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[2] M. A. Mutasodirin, R. E. Prasojo, A. F. Abka, and H. Rasyidi, “Simple hack for transformers against heavy long-text classification on a time- and memory-limited GPU service,” in Proc. 2023 10th Int. Conf. Adv. Informatics: Concept, Theory Appl. (ICAICTA), 2023, doi: 10.1109/ICAICTA59291.2023.10390269..

[3] Y. Asri, D. Kuswardani, W. N. Suliyanti, Y. O. Manullang, and A. R. Ansyari, “Sentiment Analysis Based on Indonesian Language Lexicon and IndoBERT on User Reviews PLN Mobile Application,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 38, no. 1, pp. 677–688, Apr. 2025, doi: 10.11591/ijeecs.v38.i1.pp677-688.

[4] D. Wardhani, R. Astuti, and D. D. Saputra, “Optimasi Feature Selection Text Mining: Stemming dan Stopword untuk Sentimen Analisis Aplikasi SatuSehat,” Innovative: Journal of Social Science Research, vol. 4, no. 1, pp. 7537–7548, 2024, doi: 10.31004/innovative.v4i1.8759.

[5] Y. Fauziah, B. Yuwono, and A. S. Aribowo, “Lexicon Based Sentiment Analysis in Indonesia Languages: A Systematic Literature Review,” in RSF Conference Series: Engineering and Technology, vol. 1, no. 1, 2021, pp. 363–367, doi: 10.31098/cset.v1i1.397.

[6] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Bayupati, “Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT,” Jurnal Buana Informatika, vol. 14, no. 2, 2023, doi: 10.24002/jbi.v14i02.7244.”

[7] B. Kurniawan, A. A. Aldino, and A. R. Isnain, “Sentimen analisis terhadap kebijakan penyelenggara sistem elektronik (PSE) menggunakan algoritma Bidirectional Encoder Representations from Transformers (BERT),” Jurnal Teknologi dan Sistem Informasi, vol. 3, no. 4, pp. 98–106, 2022, doi: 10.33365/jtsi.v3i4.2204.

[8] A. Faadilah, “Analisis sentimen pada ulasan aplikasi Tokopedia di Google Play Store menggunakan metode Long Short Term Memory,” Skripsi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Jakarta, Indonesia, 2020. [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/50432.

[9] A. R. Hanum, I. A. Zetha, J. N. Fajrina, R. A. Wulandari, S. C. Putri, S. P. Andina, and N. Yudistira, “Analisis kinerja algoritma klasifikasi teks BERT dalam mendeteksi berita hoaks,” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik.2024118093.

[10] F. Horn, L. Arras, G. Montavon, K.-R. Müller, and W. Samek, “Exploring text datasets by visualizing relevant words,” arXiv preprint arXiv:1707.05261, Jul. 2017. [Online]. Available: https://arxiv.org/abs/1707.05261.