

Evaluasi Kepuasan Pengguna Quick Commerce Astro Berdasarkan Ulasan Google Play Store Menggunakan Pendekatan IndoBERT

Yesi Puspita Dewi¹, Wulandari^{2*}, Mohamad Arif³

^{1,2}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

^{1,2}Ciledug Raya, RT.10/RW.2, Petukangan Utara, Kec. Pesanggrahan, DKI Jakarta 12260

³Teknik Komputer, Politeknik Prasetiya Mandiri

³Jl. Binamarga No.19, RT.06/RW.08, Baranangsiang, Kec. Bogor Timur, Kota Bogor, Jawa Barat 16143

E-mail: 'yesi.puspitadewi@budiluhur.ac.id, ^{2*}wulandari@budiluhur.ac.id, ³arif.dosen@prasetyamandiri.ac.id

(*: corresponding author)

Abstrak— Pesatnya pertumbuhan industri *quick commerce* di Indonesia, khususnya layanan Astro yang menawarkan pengiriman kebutuhan pokok instan, menuntut pemantauan kepuasan pengguna secara masif dan berkala. Sumber evaluasi utama yang kaya akan informasi adalah ulasan pengguna di platform *Google Play Store*. Namun, karakteristik ulasan yang sarat akan bahasa informal, singkatan, dan struktur semantik lokal yang kompleks menjadi tantangan tersendiri dalam analisis sentimen otomatis. Penelitian ini bertujuan untuk mengevaluasi kepuasan pengguna aplikasi Astro melalui analisis sentimen ulasan biner menggunakan pendekatan *Deep Learning* berbasis arsitektur Transformer IndoBERT (indobert-base-p2). Tahapan penelitian meliputi *scraping* data, *preprocessing*, tokenisasi berbasis algoritma WordPiece untuk mengatasi masalah *Out-of-Vocabulary* (OOV), serta proses *fine-tuning* model selama 3 epoch. Hasil pengujian kuantitatif pada 800 data uji terisolasi menunjukkan performa model yang sangat optimal, dengan capaian nilai *Accuracy* sebesar 97,75%, *Precision* 98,47%, *Recall* 99,02%, dan *F1-Score* sebesar 98,74%. Pengujian kualitatif pada data baru (*unseen data*) membuktikan ketangguhan model dalam menangkap aspek kebahasaan kontekstual, seperti ulasan berbentuk saran bernada positif dan kritik halus bernada negatif. Hasil penelitian ini membuktikan bahwa arsitektur IndoBERT hasil *fine-tuning* sangat andal digunakan sebagai instrumen evaluasi kepuasan pelanggan otomatis pada industri *quick commerce* di Indonesia.

Kata Kunci: Analisis Sentimen, Astro, Google Play Store, IndoBERT, Quick Commerce

Abstract— The rapid growth of the *quick commerce* industry in Indonesia, particularly the Astro service that offers instant delivery of basic necessities, demands massive and regular monitoring of user satisfaction. The primary source of information-rich evaluation is user reviews on the *Google Play Store* platform. However, the characteristics of reviews that are rich in informal language, abbreviations, and complex local semantic structures pose a challenge in automated sentiment analysis. This study aims to evaluate user satisfaction of the Astro application through sentiment analysis of binary reviews using a *Deep Learning* approach based on the IndoBERT Transformer architecture (indobert-base-p2). The research stages include data *scraping*, *preprocessing*, tokenization based on the WordPiece algorithm to overcome the *Out-of-Vocabulary* (OOV) problem, and a model *fine-tuning* process for 3 epochs. The results of quantitative testing on 800 isolated test data

show very optimal model performance, with an *Accuracy* value of 97.75%, *Precision* 98.47%, *Recall* 99.02%, and *F1-Score* of 98.74%. Qualitative testing on unseen data demonstrated the model's robustness in capturing contextual linguistic aspects, such as positive suggestions and subtle negative criticism. This study demonstrates that the fine-tuned IndoBERT architecture is highly reliable as an automated customer satisfaction evaluation tool in the Indonesian *quick-commerce* industry.

Keywords— Sentiment Analysis, Astro, Google Play Store, IndoBERT, Quick Commerce.

I. PENDAHULUAN

Transformasi digital di sektor ritel telah melahirkan fenomena *quick commerce* [1], sebuah evolusi dari e-commerce tradisional yang menekankan pada kecepatan pengiriman ekstrim, biasanya di bawah 30 menit. Di Indonesia, Astro telah menjadi standar baru dalam pengiriman kebutuhan pokok instan dengan janji layanan “15 menit sampai” ke tangan konsumen [2]. Kehadiran layanan ini merespons perubahan perilaku konsumen pasca-pandemi yang menginginkan efisiensi waktu dan kemudahan akses terhadap produk tanpa harus meninggalkan rumah [3]. Sebagai aplikasi yang telah diunduh lebih dari satu juta kali, Astro menghadapi tantangan besar dalam mempertahankan kualitas layanan seiring dengan pertumbuhan basis pengguna yang masif [4].

Model bisnis dengan ekspektasi kecepatan tinggi ini menciptakan tekanan operasional yang besar. Setiap interaksi pelanggan, mulai dari pemilihan item di aplikasi hingga momen penerimaan barang oleh kurir, menjadi titik evaluasi yang krusial [2][5]. Dinamika kepuasan pelanggan ini terekam secara masif dalam bentuk ulasan di platform digital seperti *Google Play Store*. Ulasan-ulasan tersebut merupakan sumber data berharga bagi perusahaan untuk mengevaluasi kualitas layanan, mulai dari kecepatan pengiriman, kesegaran produk, hingga antarmuka aplikasi [6]. Namun, dengan volume data yang mencapai ribuan setiap harinya, pemrosesan secara manual atas ribuan pasokan data setiap harinya tidak hanya memakan waktu, tetapi juga rentan terhadap penilaian yang subjektif.

Untuk mengatasi ledakan data tekstual tersebut, pendekatan Natural Language Processing (NLP) melalui analisis sentimen menjadi solusi komputasional yang krusial dalam mengekstrak polaritas opini konsumen secara otomatis [7][8][9]. Melalui klasifikasi sentimen, manajemen quick commerce dapat memetakan area operasional spesifik yang membutuhkan intervensi perbaikan, seperti manajemen rantai pasok instan maupun performa aplikasi [10][11].

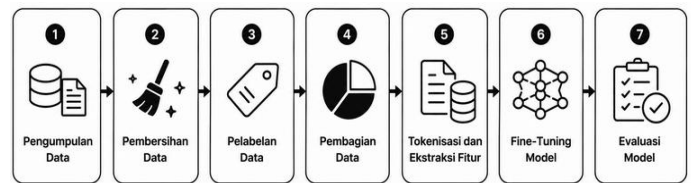
Dalam implementasinya, tantangan utama pemrosesan teks pada ulasan Play Store adalah strukturnya yang sarat akan bahasa informal, singkatan, slang, dan kalimat kontradiktif. Guna mengatasi kompleksitas semantik tersebut, teknologi Deep Learning berbasis arsitektur Transformer hadir membawa solusi mutakhir [12][7][9]. Melalui mekanisme self-attention, model Transformer mampu membaca kalimat secara dua arah (bidirectional) sehingga pemahaman konteks kontekstual menjadi lebih utuh. Keunggulan representasi bahasa ini diakomodasi dengan sangat baik oleh IndoBERT. Sebagai model bahasa pra-latih (pre-trained model), IndoBERT telah dilatih menggunakan miliaran kata dari korpus bahasa Indonesia formal dan informal, menjadikannya sangat peka terhadap nuansa linguistik lokal. Proses fine-tuning pada IndoBERT terbukti mampu mendongkrak akurasi klasifikasi teks secara signifikan pada berbagai domain digital [12][13][14].

Meski implementasi IndoBERT marak pada *e-commerce* konvensional, literatur yang mengevaluasi dinamika kepuasan pengguna di sektor quick commerce Indonesia masih sangat terbatas [6][2]. Padahal, ulasan quick commerce memiliki karakteristik unik karena konsumen sangat sensitif terhadap variabel waktu pengiriman menit-demi-menit serta kesegaran bahan pangan (perishable goods) [15][16][17]. Fenomena operasional yang dinamis ini menuntut model klasifikasi yang andal dalam memahami kompleksitas sentimen tekstual pengguna [4].

Oleh karena itu, tujuan dalam penelitian ini adalah untuk mengevaluasi kepuasan pengguna aplikasi quick commerce Astro berdasarkan ulasan Google Play Store menggunakan pendekatan IndoBERT. Kontribusi utama penelitian ini terletak pada penerapan arsitektur Transformer yang diselaraskan (fine-tuning) untuk mengenali karakteristik unik bahasa ulasan konsumen layanan instan di Indonesia. Hasil penelitian ini diharapkan dapat menyajikan metrik performa model yang komprehensif sekaligus memberikan rekomendasi strategis bagi pelaku industri quick commerce dalam mengoptimalkan layanan logistik mikro dan antarmuka digital mereka.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan klasifikasi sentimen berbasis deep learning dengan memanfaatkan model IndoBERT untuk mengidentifikasi sentimen ulasan pengguna aplikasi. Tahapan penelitian yang dilakukan terdiri dari pengumpulan data, pembersihan data, pelabelan data, pembagian data, ekstraksi fitur, optimasi model, dan prediksi akhir. Tahapan penelitian ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

Keterangan:

A. Pengumpulan Data

Dataset pada penelitian ini di dapat dari ulasan aplikasi Astro melalui teknik scraping pada Google Play Store menggunakan package Python google-play-scraper. Pengambilan data dilakukan pada tanggal 30 Mei 2026 dengan menyaring ulasan-ulasan yang paling relevan, sehingga didapatkan sampel data sebanyak 4.000 ulasan.

B. Pembersihan Data

Dataset selanjutnya dilakukan proses pembersihan (*data cleaning*) untuk meningkatkan kualitas data. Pada tahap ini dilakukan penghapusan karakter yang tidak diperlukan, seperti URL, simbol, karakter khusus, dan spasi berlebih. Penelitian ini menerapkan *light cleaning* tanpa proses stemming untuk mempertahankan konteks dan makna kata yang digunakan oleh pengguna.

C. Pelabelan Data

Pada tahap pelabelan dilakukan secara otomatis berdasarkan nilai rating yang diberikan pengguna. Rating 1 sampai 3 dikategorikan sebagai sentimen negatif (kelas 0), sedangkan rating 4 sampai 5 dikategorikan sebagai sentimen positif (kelas 1). Proses ini menghasilkan dataset berlabel yang siap digunakan dalam proses pelatihan model.

D. Pembagian Data

Dataset yang telah diberi label dibagi menjadi data latih (training data) dan data uji (testing data) menggunakan metode stratified split. Metode ini digunakan untuk mempertahankan proporsi masing-masing kelas pada kedua kelompok data. Pada penelitian ini, data dibagi dengan komposisi 80% data latih dan 20% data uji.

E. Tokenisasi dan Ekstraksi Fitur

Pada tahap ekstraksi fitur, korpus data teks dikonversi menjadi representasi vektor numerik (token IDs) yang dapat dipahami oleh model komputer menggunakan tokenizer bawaan dari model pra-latih *indobert-base-p2*. Proses tokenisasi mengimplementasikan pendekatan *subword tokenization* berbasis algoritma *WordPiece*. Mekanisme ini mampu memotong kata asing maupun kesalahan pengetikan (*typo*) menjadi sub-suku kata secara efektif, sehingga menekan risiko *Out-of-Vocabulary (OOV)* sebelum data dibungkus ke dalam arsitektur *PyTorch Tensor*.

F. Fine-Tuning Model

Model klasifikasi dibangun dengan memanfaatkan arsitektur dasar Transformer IndoBERT yang kemudian dilakukan proses penyesuaian bobot (*fine-tuning*) menggunakan data latih yang telah disiapkan. Selama proses pelatihan yang berjalan dalam beberapa *epoch*, parameter internal model diperbarui secara dinamis menggunakan fungsi

pengoptimal (*optimizer*) untuk meminimalkan nilai *loss* dan mengonvergensi bobot jaringan saraf menuju performa yang paling optimal.

G. Evaluasi Model

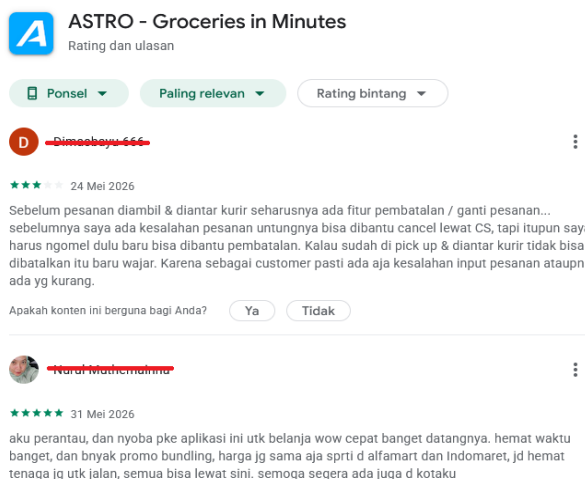
Tahap evaluasi dilakukan untuk menguji kinerja dan kapabilitas generalisasi model IndoBERT hasil *fine-tuning* melalui dua skema. Pertama, evaluasi eksperimental menggunakan data uji terisolasi (20% dataset/800 ulasan) untuk menghitung nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative* melalui *confusion matrix*. Komponen tersebut kemudian diturunkan menjadi metrik performa ilmiah meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Kedua, pengujian data baru (*unseen data*) dilakukan melalui *pipeline* inferensi otomatis untuk memprediksi polaritas ulasan riil secara langsung (*real-time*) tanpa intervensi manual. Hasil klasifikasi massal ini diekspor ke dalam berkas csv untuk mempermudah analisis tren kepuasan dan keluhan pengguna terhadap aplikasi Astro.

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari penerapan metode dalam penelitian ini didasarkan pada beberapa fase implementasi. Adapun rangkaian tahapan tersebut diuraikan sebagai berikut:

A. Scraping Data

Data pada penelitian ini berupa ulasan pengguna aplikasi Astro dari Google Play Store, proses *scraping* dilakukan pada 30 Mei 2026 menggunakan pustaka *Python google-play-scraper*. Pengambilan data difokuskan pada kategori ulasan "paling relevan" (*most relevant*), untuk memastikan kualitas informasi dan bobot umpan balik yang dianalisis. Tampilan Aplikasi Astro dapat dilihat pada gambar 1.



Gambar 2. Aplikasi Astro

Data yang didapatkan dalam proses *scraping* diambil sampel sebanyak 4.000 data ulasan mentah yang memuat teks opini, skor rating, dan waktu ulasan, selanjutnya akan ditransformasikan ke dalam tahap pra-pemrosesan teks (*text preprocessing*). Hasil *scraping* data ulasan astro ditunjukkan pada gambar 3:

	content	score
0	Sebelum pesanan diambil & diantar kurir seharu...	3
1	aku perantau, dan nyoba pke aplikasi ini utk b...	5
2	pertama kali belanja di Astro ,buat saya sanga...	5
3	TOP kalau di kota besar terbaik deh, entah aku ...	5
4	Iklannya 15 menit sampai, tapi sudah beberapa ...	1

Gambar 3. Ulasan aplikasi Astro

B. Pembersihan Data

Tahap pembersihan data (*data cleaning*) memegang peranan penting dalam tahap pra-pemrosesan (*preprocessing*) teks untuk meminimalisasi noise serta mendongkrak akurasi analisis pada model deep learning. Pada fase ini, ulasan mentah dari Google Play Store ditransformasikan ke dalam format yang lebih bersih dan standar melalui serangkaian teknik eliminasi gangguan kebahasaan. Proses ini mencakup penghapusan elemen non-alfabet yang tidak merepresentasikan ekspresi sentimen secara langsung, seperti tautan URL, simbol digital, angka, karakter khusus, serta tanda baca. Selain itu, dilakukan pula konversi seluruh teks menjadi huruf kecil (*case folding*) guna memastikan konsistensi bentuk kata, serta perapian jarak antar-kata melalui reduksi spasi berlebih (*strip spaces*). Implementasi dari seluruh rangkaian teknik pembersihan data ini disajikan secara komparatif pada gambar 4.

	content	clean_content
0	Sebelum pesanan diambil & diantar kurir seharu...	sebelum pesanan diambil diantar kurir seharusn...
1	Iklannya 15 menit sampai, tapi sudah beberapa ...	iklannya 15 menit sampai tapi sudah beberapa k...
2	Aplikasi yang sangat membantu, pesanan selalu ...	aplikasi yang sangat membantu pesanan selalu d...
3	aku baru pertama kali pakai aplikasi astro kar...	aku baru pertama kali pakai aplikasi astro kar...
4	Astro, pls, jangkauannya sampein ke Sumatra do...	astro pls jangkauannya sampein ke sumatra dong...
5	TOP kalau di kota besar terbaik deh, entah aku ...	top kalau di kota besar terbaik deh entah aku t...
6	sekarang aku belanja di astro terus, karna ter...	sekarang aku belanja di astro terus karna tern...
7	keren banget beneran secepat itu nganternya, g...	keren banget beneran secepat itu nganternya ga...
8	Gini deh. Lu yang bingung + mager juga ke pasa...	gini deh lu yang bingung mager juga ke pasar n...
9	drivernya sering banget salah kirim, padahal s...	drivernya sering banget salah kirim padahal se...

Gambar 4. Hasil *cleaning*

Dalam penelitian ini penulis sengaja menerapkan metode pembersihan ringan (*light cleaning*) dengan mengeklusi atau tidak mengikutsertakan proses pemotongan kata imbuhan (*stemming*). Hal ini dikarenakan karakteristik unik dari korpus ulasan media digital yang bersifat kasual dan sangat kontekstual. Dengan tidak mengikuti tahap *stemming*, penelitian ini dapat mempertahankan bentuk asli kosakata serta menjaga keutuhan konteks semantik dari setiap kalimat.

Selain itu, pendekatan ini krusial untuk mengonservasi penekanan makna emosional—seperti penggunaan kata seru, partikel penegas (misalnya, "pun", "deh", "dong"), maupun reduplikasi kata—yang secara alamiah diekspresikan oleh pengguna saat mengulas layanan aplikasi Astro. Penerapan konfigurasi pra-pemrosesan yang selektif ini terbukti sangat mendukung arsitektur berbasis Transformer seperti IndoBERT, yang secara inheren mengandalkan struktur kalimat utuh dan

nuansa kebahasaan lokal untuk melakukan ekstraksi fitur sentimen secara optimal[12][18].

C. Pelabelan Data

Tahap selanjutnya pelabelan data yaitu pelabelan dilakukan secara otomatis menggunakan teknik *rating-based labeling* berskala 1–5 [9][7]. Data ditransformasikan ke dalam dua kelas sentimen biner: rating 1 hingga 3 dikategorikan sebagai sentimen negatif (Kelas 0) untuk merepresentasikan keluhan pengguna, sedangkan rating 4 dan 5 diklasifikasikan sebagai sentimen positif (Kelas 1) yang menunjukkan kepuasan pengguna. Pendekatan otomatis ini menghasilkan korpus data berlabel yang konsisten dan siap digunakan untuk tahap ekstraksi fitur serta pelatihan model. Hasil pelabelan ditunjukkan pada gambar 5.

score		clean_content	label
0	3	sebelum pesanan diambil diantar kurir seharus...	0
1	1	iklannya 15 menit sampai tapi sudah beberapa k...	0
2	5	aplikasi yang sangat membantu pesanan selalu d...	1
3	5	aku baru pertama kali pakai aplikasi astro kar...	1
4	4	astro pls jangkauannya sampein ke sumatra dong...	1
5	5	top kalau di kota besar terbaik deh entah aku t...	1
6	5	sekarang aku belanja di astro terus karna tern...	1
7	5	keren banget beneran secepat itu nganternya ga...	1
8	5	gini deh lu yang bingung mager juga ke pasar n...	1
9	5	drivernya sering banget salah kirim padahal se...	1

Gambar 5. Pelabelan Data

D. Pembagian Dataset

Dataset berlabel yang telah terbentuk kemudian dieksekusi ke dalam tahap pembagian data menjadi data latih (training data) dan data uji (testing data). Penelitian ini menerapkan metode stratified split untuk memastikan bahwa proporsi distribusi kelas sentimen (positif dan negatif) tetap seimbang dan konsisten pada kedua sub-kelompok data tersebut. Komposisi pembagian data yang digunakan adalah 80% untuk data latih dan 20% untuk data uji. Dari total 4.000 data ulasan, rasio ini menghasilkan 3.200 data yang dialokasikan untuk melatih performa algoritma, serta 800 data yang diisolasi khusus untuk menguji akurasi dan generalisasi model klasifikasi yang dikembangkan. Pembagian dataset ditunjukkan pada gambar 6.

```
x = df['clean_content'].values
y = df['label'].values

train_texts, val_texts, train_labels, val_labels = train_test_split(
    x, y, test_size=0.2, random_state=42, stratify=y
)
print(f>Data Training: {len(train_texts)} | Data Validasi: {len(val_texts)}")

Data Training: 3200 | Data Validasi: 800
```

Gambar 6. Pembagian Data

E. Tokenisasi dan Ekstraksi Fitur

Tahap ekstraksi fitur menggunakan tokenizer model dasar IndoBERT (*indobert-base-p2*) berbasis algoritma WordPiece. Pendekatan subword tokenization ini efektif memecah kata berimbuhan maupun typo (seperti "disinipelayanan" menjadi disini, ##pel, ##ayanan), sehingga mampu mengatasi masalah Out-of-Vocabulary (OOV) pada ulasan media sosial. Hasil tokenisasi berupa token IDs kemudian ditransformasikan ke dalam PyTorch Tensor sebagai input pelatihan, hasil tokenizer ditunjukkan pada gambar 7.

No	Content	Token (IndoBERT)	Token IDs (Numerik)
1	sering pakai disinipelayanan cepat aplikadi si...	[CLS], sering, pakai, disini, ##pel, ##ayanan,...	[2, 884, 2468, 1612, 6509, 23792, 972, 13144, ...
2	baru ini dapet aplikasi belanja keren bgtsiny...	[CLS], baru, ini, dapet, aplikasi, belanja, ke...	[2, 440, 92, 7120, 1339, 3866, 4076, 6948, 556, ...
3	aplikasi yg sangat memudahkan mencari barang b...	[CLS], aplikasi, yg, sangat, memudahkan, menca...	[2, 1339, 741, 310, 4628, 1105, 963, 3866, 5, ...
4	ternyata begini rasanya belanja di rumah yang ...	[CLS], ternyata, begini, rasanya, belanja, di,...	[2, 1369, 6838, 3514, 3866, 26, 448, 34, 1375, ...
5	sekarang semua serba mudah astro sangat memper...	[CLS], sekarang, semua, serba, mudah, astr, ##...	[2, 747, 366, 7734, 783, 7451, 30370, 310, 737, ...

Gambar 7. Tokenization

F. Fine-Tuning Model

Model klasifikasi dibangun menggunakan arsitektur IndoBERT (*indobert-base-p2*) melalui proses fine-tuning pada data latih selama 3 epoch. Selama pelatihan, parameter model disesuaikan secara dinamis untuk meminimalkan loss dan mengoptimalkan performa. Evaluasi kemampuan generalisasi model diuji menggunakan data validasi berdasarkan metrik Accuracy, Precision, Recall, dan F1-score. Hasil ditunjukkan pada gambar 8 yang menunjukkan tingkat konvergensi yang sangat baik.

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0.147274	0.095478	0.972500	0.984658	0.980556	0.988796
2	0.045058	0.115537	0.975000	0.986034	0.983287	0.988796
3	0.030780	0.134158	0.977500	0.987430	0.984680	0.990196

Gambar 8. Hasil Fine-Tuning

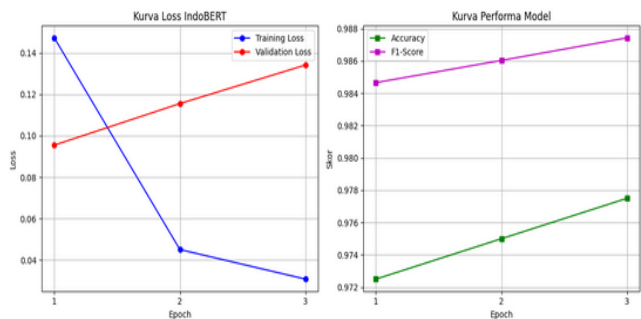
Nilai Training Loss menurun drastis dari 0,1473 (epoch 1) menjadi 0,0308 (epoch 3), menandakan model mampu mempelajari pola fitur teks secara optimal. Penurunan error ini berbanding lurus dengan peningkatan metrik evaluasi secara berkala. Pada epoch akhir, model mencapai performa puncak dengan Akurasi 97,75%, Precision 98,47%, Recall 99,02%, dan F1-score 98,74%. Meskipun Validation Loss meningkat tipis dari 0,0955 ke 0,1342, stabilitas metrik utama (Akurasi dan F1-score) yang tetap meningkat membuktikan bahwa arsitektur IndoBERT sangat andal dalam mengklasifikasikan sentimen ulasan biner aplikasi Astro tanpa mengalami *overfitting*.

G. Evaluasi Model

Pada tahap evaluasi dilakukan dua pengujian, pertama pengujian performa arsitektur IndoBERT hasil *fine-tuning* dilakukan melalui dua skema, yaitu evaluasi kuantitatif

menggunakan data uji terisolasi (20% dataset) dan pengujian kualitatif pada data ulasan baru (*unseen data*).

Proses konvergensi dan stabilitas internal model dipetakan secara berkala pada setiap *epoch* melalui indikator *loss* dan skor performa yang disajikan pada gambar 9.

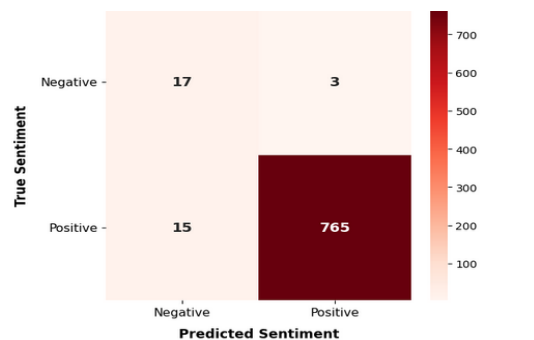


Gambar 9. Kurva Pelatihan Model IndoBERT: (a) Kurva Loss, (b) Kurva Performa Metrik

Berdasarkan Kurva Loss IndoBERT pada gambar 9(kiri), nilai *Training Loss* menunjukkan tren penurunan yang sangat signifikan, bergerak dari 0,147 pada *epoch* 1 hingga mencapai titik terendah sebesar 0,031 pada *epoch* 3, yang mengindikasikan bahwa model mampu mempelajari fitur semantik dari korpus ulasan Astro secara agresif. Meskipun demikian, kurva *Validation Loss* menunjukkan tren peningkatan yang linier, bergerak dari 0,095 ke 0,134, dimana fenomena divergensi ini mengindikasikan adanya gejala *overfitting* ringan seiring bertambahnya *epoch* karena model mulai terspesialisasi pada data latih, sehingga *epoch* awal (antara *epoch* 1 dan 2) dapat diidentifikasi sebagai titik *stopping* optimal untuk generalisasi terbaik. Kendati terdapat kenaikan pada *validation loss*.

Kapasitas klasifikasi model secara keseluruhan tetap berada pada kategori superior sebagaimana ditunjukkan pada Kurva Performa Model di gambar 9(kanan)), dengan nilai *Accuracy* dan *F1-Score* yang mengalami peningkatan konsisten secara simultan di setiap fasenya, sehingga stabilitas grafik performa ini memberikan fondasi yang kuat bahwa bobot parameter yang diperbarui selama proses *fine-tuning* berhasil mentransfer pengetahuan bahasa alami secara optimal.

Selanjutnya, untuk membuktikan performa kuantitatif tersebut secara riil, pengujian dilakukan menggunakan *confusion matrix* terhadap 800 data uji terisolasi yang disajikan pada gambar 10



Gambar 10. Hasil Confusion Matrix

Model terbukti memiliki akurasi dan presisi yang sangat tinggi. Model berhasil mengklasifikasikan 765 data sebagai *True Positive* dan 17 data sebagai *True Negative*. Sebaliknya, margin kesalahan model sangat minim, dengan hanya 3 data *False Positive* dan 15 data *False Negative*. Dominasi data pada diagonal utama matriks ini (782 prediksi benar) mempertegas ketangguhan model dalam meminimalkan bias, dengan capaian nilai *Accuracy* sebesar 97,75% dan *F1-Score* sebesar 98,74%.

Pengujian kedua dengan menguji menggunakan data baru,, pengujian tersebut ditunjukkan pada gambar 11.

```
ulasan_baru = [
    "Barang sampai dalam kondisi segar bgt, kurirnya juga ramah. Pertahankan kualitasnya",
    "Aplikasi jelek banget sering force close sendiri pas mau checkout belanjaan.",
    "Bintang 3 dulu deh, barangnya oke tapi pengirimannya tidak secepat yang diiklankan.",
    "Diskonannya tolong ditambah lagi dong biar makin langganan.",
    "Sangat kecewa belanja di sini, sayurnya ada yang busuk dan CS nya susah dihubungi."
]

print(f"Mempersiapkan {len(ulasan_baru)} data ulasan baru...")

# 4. Melakukan Prediksi Massal (Inference)
hasil_prediksi = []

for teks in ulasan_baru:
    teks_bersih = clean_text_for_bert(teks)

    # Tokenisasi teks input
    inputs = tokenizer(teks_bersih, return_tensors="pt", truncation=True, padding=True,
    inputs = {k: v.to(device) for k, v in inputs.items()})

    # Prediksi menggunakan model
    with torch.no_grad():
        outputs = model(**inputs)

    # Ambil kelas dengan probabilitas tertinggi
    id_kelas = torch.argmax(outputs.logits, dim=-1).item()
    sentimen = "POSITIF" if id_kelas == 1 else "NEGATIF"

    hasil_prediksi.append({
        'Ulasan Mentah': teks,
        'Sentimen Prediksi': sentimen
    })
```

Gambar 11. Ulasan Data Baru

Dari Ulasan baru diatas didapatkan hasil sebagai berikut:

Ulasan Mentah	Sentimen Prediksi
Barang sampai dalam kondisi segar bgt, kurirnya juga ramah. Pertahankan kualitasnya Astro!	POSITIF
Aplikasi jelek banget sering force close sendiri pas mau checkout belanjaan.	NEGATIF
Bintang 3 dulu deh, barangnya oke tapi pengirimannya tidak secepat yang diiklankan.	NEGATIF
Diskonannya tolong ditambah lagi dong biar makin langganan.	POSITIF
Sangat kecewa belanja di sini, sayurnya ada yang busuk dan CS nya susah dihubungi.	NEGATIF

Gambar 12. Hasil pengujian Data Baru

Berdasarkan gambar 12 diatas, hasil pengujian kualitatif pada data ulasan baru (*unseen data*) menunjukkan bahwa model IndoBERT memiliki pemahaman kontekstual yang

sangat baik. Model tidak hanya mengenali polaritas eksplisit (seperti kata "*segar bgt*" sebagai POSITIF dan "*sangat kecewa*" sebagai NEGATIF), tetapi juga adaptif terhadap struktur kalimat yang lebih kompleks. Kemampuan ini terlihat pada akurasi prediksi POSITIF untuk ulasan berupa saran ("*Diskonannya tolong ditambah...*") serta prediksi NEGATIF untuk kritik halus ("*Bintang 3 dulu deh...*"). Hal ini membuktikan bahwa proses *fine-tuning* berhasil membuat model menangkap batasan semantik dan nuansa bahasa lokal secara optimal.

IV. PENUTUP

Berdasarkan hasil penelitian dan pembahasan yang telah dipaparkan, dapat disimpulkan bahwa pendekatan Deep Learning menggunakan arsitektur Transformer IndoBERT (indobert-base-p2) terbukti sangat andal dalam mengevaluasi kepuasan pengguna aplikasi quick commerce Astro berdasarkan ulasan di Google Play Store. Melalui proses fine-tuning selama 3 epoch, model menunjukkan tingkat konvergensi yang sangat baik dengan penurunan nilai training loss yang drastis dari 0,1473 menjadi 0,0308. Secara kuantitatif, pengujian model pada 800 data uji terisolasi menghasilkan performa puncak yang sangat impresif, ditandai dengan nilai Accuracy sebesar 97,75% dan F1-Score sebesar 98,74% (di mana model sukses memprediksi dengan benar 765 sentimen positif dan 17 sentimen negatif).

Secara kualitatif, pengujian pada data baru menegaskan kapabilitas model dalam memahami nuansa bahasa natural konsumen lokal secara kontekstual, termasuk sensitivitas dalam membedakan antara ulasan bersyarat/saran yang bernada loyalitas (positif) dengan ekspresi kritik halus maupun sarkasme ringan (negatif). Secara keseluruhan, sistem klasifikasi ini berhasil memetakan bahwa mayoritas pengguna merasa puas terhadap aspek pengiriman instan dan kualitas produk Astro, sekaligus mendeteksi area minor yang memerlukan perbaikan operasional logistik atau antarmuka aplikasi. Untuk penelitian selanjutnya, disarankan agar melakukan klasifikasi sentimen berbasis aspek (Aspect-Based Sentiment Analysis) guna mengelompokkan keluhan konsumen secara lebih spesifik serta menguji model pada skema multi-kelas (positif, netral, negatif).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih dan apresiasi yang sebesar-besarnya kepada rekan-rekan sejawat di lingkungan akademis serta seluruh pihak terkait yang telah memberikan kontribusi pemikiran, validasi metodologi, serta diskusi yang konstruktif selama proses penelitian ini. Dukungan, saran, dan kritik membangun yang diberikan oleh berbagai pihak tersebut sangat berharga dalam menyempurnakan analisis hingga penelitian mengenai kepuasan pengguna aplikasi Astro ini dapat diselesaikan dengan baik dan mencapai standar publikasi ilmiah yang diharapkan.

REFERENSI

- [1] N. S. Dissurul *et al.*, "Transformasi Pola Belanja Konsumen dari Ritel Konvensional Menuju Quick Commerce Melalui Layanan Pengiriman Instan," *JUBIKIN: Jurnal Bisnis Kreatif dan*

- Inovatif*, vol. 2, no. pp. 241–256, 2025, doi: 10.61132/jubikin.v2i4.1138.
- [2] F. Titania and M. G. Yoedjadi, "Efektivitas Persepsi Kegunaan dan Tagline '15 Menit Sampai' terhadap Keputusan Pembelian Layanan Astro E-grocery," *Prologia*, vol. 10, N, pp. 167–176, 2026, doi: 10.24912/pr.v10i1.36347.
- [3] J. Torang, M. Simamora, and F. Muslim, "Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Pembelian Melalui Online Shop Pada Mahasiswa Pendidikan Ekonomi Angkatan 2020 Universitas Jambi," *JOTE (JOURNAL Teach. Educ.)*, vol. 5, no. 4, pp. 98–107, 2024, doi: 10.31004/jote.v5i4.29836.
- [4] S. A. Prasetyo and W. T. Atmojo, "Analisis Sentimen Terhadap Penggunaan Aplikasi Astro Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, vol. 7, no. April 2021, pp. 2149–2158, 2023, doi: 10.30865/mib.v7i4.6750.
- [5] M. Anisa and M. S. Purwanegara, "Entrepreneurship Bisnis Manajemen Akuntansi Unraveling the rise and fall of quick commerce in Indonesia: insights from industry experts," *E-BISMA*, vol. 6, no. 1, pp. 19–31, 2025, doi: 10.37631/ebisma.v6i1.1753.
- [6] A. K. Ibrahim, M. Mustikasari, and I. Bastian, "Analisis Sentimen Pada Ulasan Aplikasi Astro – Groceries In Minutes Di Google Play Store Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT)," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.6033.
- [7] K. C. Pradhisa and R. Fajriyah, "Analisis Sentimen Ulasan Pengguna E-commerce di Google Play Store Menggunakan Metode IndoBERT," *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 92–104, 2024, doi: 10.47065/bits.v6i1.5247.
- [8] I. Mursidah, *et al.*, "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT," *J. Minfo Polgan*, vol. 14, pp. 3349–3359, 2025, doi: 10.33395/jmp.v14i2.15751.
- [9] R. Nugroho, N. Azka, W. Sayudha, and R. Graha, "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 495–505, 2025, doi: 10.35870/jtik.v9i2.3340.
- [10] A. R. Putra and D. E. Ratnawati, "Analisis Sentimen Berbasis Aspek Pada Aplikasi Mobile Menggunakan Naïve Bayes Berdasarkan Ulasan Pengguna Playstore (Studi Kasus : Jconnect Mobile)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 293–300, 2025, doi: 10.25126/jtiik.2025127556.
- [11] M. Faiq, P. Lestari, L. Belluano, and H. Darwis, "Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest," *LINIER*, vol. 2, no. 3, pp. 380–388, 2025, doi: 10.33096/linier.v2i3.3149.
- [12] F. Koto, A. Rahimi, J. L. Han, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *arxiv logo*, 2020, doi: 10.48550/arXiv.2011.00677.
- [13] A. L. Maysara, "Optimalisasi Akurasi dan Stabilitas Analisis Sentimen Ulasan E-Commerce Indonesia melalui Fine-Tuning Transformer IndoBERT," *Infotekmesin*, vol. 17, no. 01, pp. 9–17, 2026, doi: 10.35970/infotekmesin.v17i1.3037.
- [14] L. N. Hakim *et al.*, "Klasifikasi Komentar Toksik Berbahasa Indonesia di Media Sosial Berbasis Fine-Tuning IndoBERT," *JURIKOM (Jurnal Ris. Komputer)*, vol. 13, no. 1, pp. 202–209, 2026, doi: 10.30865/jurikom.v13i1.9449.
- [15] V. S. Teotia, "Impact of Delivery Speed on Customer Satisfaction in Online Shopping: A Focus on Quick Commerce," *Int. J. Sci. Res. Eng. Manag.*, pp. 8–12, 2025, doi: 10.55041/IJSREM49149.

- [16] A. Harter, L. Stich, and M. Spann, "The Effect of Delivery Time on Repurchase Behavior in Quick Commerce," *Sage Journals*, vol. 28, no. 2, pp. 211–227, 2025, doi: 10.1177/10946705241236961.
- [17] A. M. Putri, W. K. Nofa, and D. A. P. Hapsari, "Penerapan Metode Bert untuk Analisis Sentimen Ulasan Pengguna Aplikasi Segari Di Google Play Store," vol. 4, no. 1, pp. 89–104, 2025, doi: 10.56127/juit.v4i1.1902.
- [18] M. Akmal *et al.*, "Dampak Pra-pemrosesan Teks pada Akurasi Analisis Sentimen Multi-Aspek Menggunakan IndoBERT," *Sains Data Jurnal Studi Matematika dan Teknologi*, vol. 4, no. 1, pp. 42–48, 2026, doi: 10.52620/sainsdata.v4i1.301.