

ANALISIS SENTIMEN PENJUALAN SMARTPHONE PADA TOKOPEDIA MENGGUNAKAN INDOBERT

SKRIPSI

Oleh:

CALVIN
NIM. 221130106
DEVIN LEONARDO
NIM. 221130004
JEFFREY STANLIEM
NIM. 221130072



PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026

SENTIMENT ANALYSIS OF SMARTPHONE SALES ON TOKOPEDIA USING INDOBERT

UNDERGRADUATE THESIS

By:

CALVIN

ID NUMBER. 221130106

DEVIN LEONARDO

ID NUMBER. 221130004

JEFFREY STANLIEM

ID NUMBER. 221130072



**UNDERGRADUATE PROGRAM OF INFORMATION TECHNOLOGY
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

**ANALISIS SENTIMEN PENJUALAN SMARTPHONE PADA
TOKOPEDIA MENGGUNAKAN INDOBERT**

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknologi Informasi

Oleh:

CALVIN
NIM. 221130106
DEVIN LEONARDO
NIM. 221130004
JEFFREY STANLIEM
NIM. 221130072

Disetujui Oleh:

Dosen Pembimbing I,



Wulan Sri Lestari, S.Kom., M.Kom.

Dosen Pembimbing II,



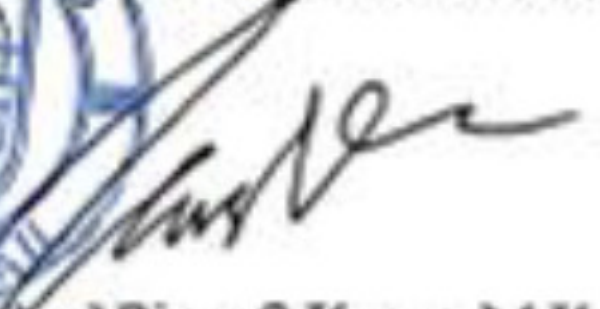
Apriyanto Haldi, S.Kom., M.Kom.

Medan, 27 Juli 2026

Diketahui dan Disahkan Oleh:



Ketua Program Studi
Teknologi Informasi,


Muska Ulina, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Calvin

NIM : 221130106

Saya telah melaksanakan penelitian dan penulisan Tugas Akhir dengan judul dan tempat penelitian sebagai berikut:

Judul Tugas Akhir : Analisis Sentimen Penjualan Smartphone Pada Tokopedia Menggunakan IndoBERT

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Tugas Akhir tersebut, dengan ini saya menyatakan dengan sebenarnya bahwa penelitian dan penulisan Tugas Akhir tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tugas Akhir saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tugas Akhir saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 28 Juni 2026

Saya yang membuat pernyataan,



Calvin

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Devin Leonardo

NIM : 221130004

Saya telah melaksanakan penelitian dan penulisan Tugas Akhir dengan judul dan tempat penelitian sebagai berikut:

Judul Tugas Akhir : Analisis Sentimen Penjualan Smartphone Pada Tokopedia Menggunakan IndoBERT

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Tugas Akhir tersebut, dengan ini saya menyatakan dengan sebenarnya bahwa penelitian dan penulisan Tugas Akhir tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tugas Akhir saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tugas Akhir saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 28 Juni 2026

Saya yang membuat pernyataan,



Devin Leonardo

ANALISIS SENTIMEN PENJUALAN SMARTPHONE PADA TOKOPEDIA MENGGUNAKAN INDOBERT

Abstrak

Peningkatan signifikan pasar smartphone di Indonesia, yang tercatat terus bertumbuh mendorong pesatnya aktivitas transaksi pada platform e-commerce seperti Tokopedia. Dalam e-commerce, ulasan pembeli menjadi salah satu unsur penting yang mencerminkan pengalaman dan preferensi pembeli agar penjual dapat meningkatkan kualitas layanannya. Namun, volume data yang terus bertambah serta penggunaan bahasa informal dengan singkatan menjadi tantangan tersendiri dalam proses klasifikasi ulasan secara manual, yang berpotensi menyebabkan hilangnya informasi berharga bagi penjual. Penelitian ini bertujuan untuk mengimplementasikan analisis sentimen guna mengklasifikasikan ulasan pembeli secara efisien menggunakan model IndoBERT. Ulasan pembeli smartphone yang dikumpulkan berada dalam rentang waktu Januari 2024 sampai Mei 2026, didapatkan sebanyak 8.203 baris data yang seimbang antara kelas positif dan negatif. Dataset melalui tahapan preprocessing terlebih dahulu sebelum digunakan untuk melatih model dengan rasio pembagian data 80:10:10. Berdasarkan hasil pengujian 30 kombinasi hyperparameter menggunakan kurva training dan validation loss, confusion matrix dan classification report, performa model yang paling optimal menunjukkan nilai accuracy sebesar 93%, precision sebesar 93%, recall sebesar 92% dan F1-score sebesar 93%. Tingginya akurasi yang dihasilkan oleh model IndoBERT menunjukkan keberhasilan model dalam membedakan ulasan positif dan negatif sehingga model yang dihasilkan dapat digunakan untuk tugas pengklasifikasian ulasan pembeli smartphone yang lebih efektif dan efisien.

Kata kunci: Analisis Sentimen, IndoBERT, E-commerce, Smartphone

Abstract

The significant growth of the smartphone market in Indonesia, which continues to expand, has driven a rapid increase in transaction activities on e-commerce platforms such as Tokopedia. In e-commerce, customer reviews are an essential element that reflect buyer experiences and preferences, enabling sellers to improve service quality. However, the growing volume of data and the use of informal language with abbreviations pose challenges for manual review classification, potentially leading to the loss of valuable information for sellers. This study aims to implement sentiment analysis to efficiently classify customer reviews using the IndoBERT model. The smartphone reviews were collected between January 2024 and May 2026, totaling 8,203 data entries balanced between positive and negative classes. The dataset underwent preprocessing before being used to train the model with an 80:10:10 data split ratio. Based on testing 30 hyperparameter combinations using training and validation loss curves, confusion matrix, and classification report, the optimal model achieved 93% accuracy, 93% precision, 92% recall, and 93% F1-score. The high accuracy indicates the model's effectiveness in distinguishing positive and negative reviews, making it suitable for more effective and efficient customer review classification.

Keywords: Sentiment Analysis, IndoBERT, E-commerce, Smartphone

KATA PENGANTAR

Puji Syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas berkat dan anugerah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“ANALISIS SENTIMEN PENJUALAN SMARTPHONE PADA TOKOPEDIA MENGGUNAKAN INDOBERT”**.

Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Program Studi Teknologi Informasi, Fakultas Informatika, Universitas Mikroskil.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada masa penyusunan skripsi ini sangatlah sulit bagi penulis untuk menyelesaikannya. Oleh karena itu, penulis mengucapkan terima kasih kepada:

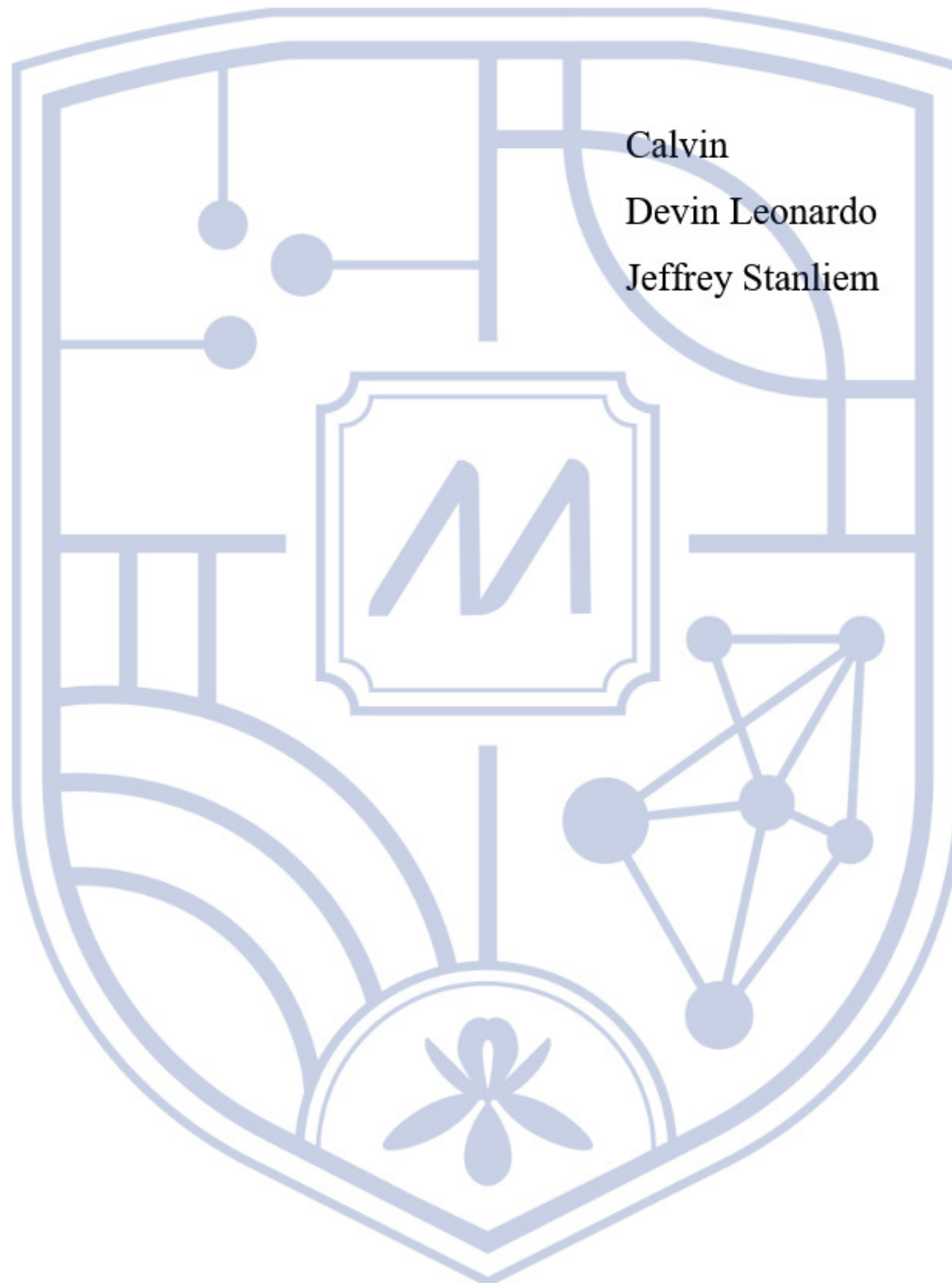
1. Ibu Wulan Sri Lestari, S.Kom., M.Kom., selaku Dosen Pembimbing I, atas bimbingan, arahan, serta dukungan yang diberikan selama proses penyusunan Tugas Akhir ini.
2. Bapak Apriyanto Halim, S.Kom., M.Kom., selaku Dosen Pembimbing II, atas bimbingan, arahan, serta dukungan yang diberikan selama proses penyusunan Tugas Akhir ini.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan, atas dukungan dan perhatian yang diberikan selama penulis menempuh studi.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan, atas dukungan dan perhatian yang diberikan selama penulis menempuh studi.
5. Ibu Mustika Ulina, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknologi Informasi, Fakultas Informatika Universitas Mikroskil Medan, atas dukungan dan perhatian selama penulis menempuh studi.
6. Bapak dan Ibu Dosen Universitas Mikroskil yang telah memberikan ilmunya selama penulis menempuh pendidikan di Universitas Mikroskil Medan.
7. Orang tua dan keluarga tercinta semua penulis, atas doa, semangat, dan dukungan yang tiada henti dalam setiap langkah peneliti.
8. Rekan-rekan serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan dukungan selama proses penyusunan Tugas Akhir ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan dikarenakan keterbatasan pengalaman dan pengetahuan yang penulis miliki. Oleh karena itu, penulis

mengharapkan segala bentuk saran serta masukan, bahkan kritik yang membangun dari berbagai pihak. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan bagi para pembaca. Terima Kasih.

Medan, 28 Juni 2026

Penulis,



DAFTAR ISI

Abstrak	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR	5
2.1 <i>Crawling</i> Data	5
2.2 <i>Data Preprocessing</i>	6
2.3 Pelabelan Data	6
2.4 <i>Data Mining</i>	8
2.5 <i>Natural Language Processing</i>	9
2.6 BERT	10
2.6.1 IndoBERT	11
2.6.2 <i>Attention Mechanism</i>	13
2.6.3 <i>Fine-Tuning</i> IndoBERT	14
2.7 Analisis Sentimen	18
2.8 <i>E-Commerce</i>	20
2.9 Tokopedia	21
2.10 Penjualan <i>Smartphone</i>	22
2.11 <i>Confusion Matrix</i>	23
BAB III TAHAPAN PELAKSANAAN	25
3.1 Pengumpulan Data	26
3.2 Analisis Proses	27
3.2.1 Pelabelan Data	27

3.2.2 <i>Preprocessing Data</i>	29
3.2.3 <i>Data Splitting</i>	33
3.2.4 Model Analisis Sentimen IndoBERT	34
3.2.5 Evaluasi Model.....	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil	45
4.1.1 Hasil Pengumpulan <i>Dataset</i>	45
4.1.2 Hasil <i>Preprocessing Dataset</i>	47
4.1.3 Hasil <i>Training</i> Model IndoBERT	55
4.1.3 Hasil Evaluasi Model IndoBERT	60
4.2 Pembahasan	64
BAB V PENUTUP	66
5.1 Kesimpulan	66
5.2 Saran	66
DAFTAR PUSTAKA	68
DAFTAR RIWAYAT HIDUP	101

DAFTAR GAMBAR

Gambar 2.1 <i>Framework Scrappy</i> Berbasis Python [16].....	4
Gambar 2.2 Arsitektur BERT [32]	10
Gambar 2.3 Alur Kerja IndoBERT [39].....	11
Gambar 2.4 PT. GoTo Gojek Tokopedia Tbk. [80]	21
Gambar 2.5 Total Penjualan <i>Smartphone</i> di Dunia (Q1 2023 – Q3 2025) [1, 83].....	22
Gambar 3.1 <i>Flowchart</i> Analisis Proses	26
Gambar 3.2 Tahapan <i>Preprocessing Data</i>	28
Gambar 3.3 Tahapan Membangun Model Analisis Sentimen IndoBERT	33
Gambar 3.4 <i>Embedding</i> IndoBERT [38].....	34
Gambar 4.1 Kode program pengumpulan <i>dataset</i> (sambungan).....	45
Gambar 4.2 Kode program tahapan <i>data cleaning</i>	47
Gambar 4.3 Kode program tahapan <i>case folding</i>	48
Gambar 4.4 Kode program tahapan <i>tokenization</i>	49
Gambar 4.5 Kode program tahapan <i>normalization</i>	50
Gambar 4.6 Kode program tahapan <i>stopword removal</i>	52
Gambar 4.7 Kode program tahapan <i>lemmatization</i>	53
Gambar 4.8 Kode program <i>data splitting</i> dan <i>fine-tuning</i> model (sambungan).....	56
Gambar 4.9 Grafik <i>loss</i> pada 6 model dengan akurasi terbaik.....	58
Gambar 4.10 Grafik <i>loss</i> pada 6 model dengan akurasi terbaik pada dataset kedua.....	60
Gambar 4.11 <i>Confusion matrix</i> dataset kedua eksperimen #2	61
Gambar 4.12 <i>Classification report</i> dataset kedua eksperimen #2	62
Gambar 4.13 Hasil Analisis Polaritas Sentimen Sampel.....	65

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> [87]	22
Tabel 3.1 Sampel <i>Dataset</i> Hasil <i>Web Crawling</i>	25
Tabel 3.2 Aturan Pelabelan Kelas	27
Tabel 3.3 Sampel Hasil Pelabelan Manual	27
Tabel 3.4 Tahapan <i>Data Cleaning</i>	29
Tabel 3.5 Tahapan <i>Case Folding</i>	29
Tabel 3.6 Tahapan <i>Tokenization</i>	30
Tabel 3.7 Tahapan <i>Normalization</i>	30
Tabel 3.8 Tahapan <i>Stopword Removal</i>	31
Tabel 3.9 Tahapan <i>Lemmatization</i>	32
Tabel 3.10 Pembagian Data Latih, Validasi dan Uji	33
Tabel 3.11 Hasil Tokenisasi <i>WordPiece</i> dan IDs IndoBERT	34
Tabel 3.12 Nilai Q, K, V dalam 1 <i>head</i> model IndoBERT	37
Tabel 3.13 Nilai <i>Multi-head Self-attention</i>	38
Tabel 3.14 Nilai <i>Logits</i> Teks	39
Tabel 4.1 Sampel Hasil Kumpulan <i>Dataset</i>	46
Tabel 4.2 Hasil tahapan <i>data cleaning</i>	47
Tabel 4.3 Hasil tahapan <i>case folding</i>	48
Tabel 4.4 Hasil tahapan <i>tokenization</i>	49
Tabel 4.5 Hasil tahapan <i>normalization</i>	51
Tabel 4.6 Hasil tahapan <i>stopword removal</i>	52
Tabel 4.7 Hasil tahapan <i>lemmatization</i>	53
Tabel 4.8 Kombinasi nilai <i>hyperparameter</i> yang diuji	54
Tabel 4.9 Hasil pengujian nilai <i>hyperparameter</i>	56
Tabel 4.10 Hasil Pengujian <i>Dataset</i> Kedua	59
Tabel 4.11 Sampel Hasil TP dan TN Model	62
Tabel 4.12 Sampel Hasil FP dan FN Model	64

DAFTAR LAMPIRAN

Lampiran 1 Gambar.....	75
Lampiran 2 Tabel.....	84

