

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI TOKOPEDIA MENGGUNAKAN MODEL BILSTM

TESIS

Oleh:

**ERIK WARDI
NIM. 231231046**



**PROGRAM STUDI S-2 TEKNOLOGI INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2025**

SENTIMENT ANALYSIS OF USER REVIEWS OF TOKOPEDIA APPLICATION USING BILSTM MODEL

THESIS

By:

**ERIK WARDI
ID NUMBER. 231231046**



**MAJOR OF S-2 INFORMATION TECHNOLOGY
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2025**

LEMBARAN PENGESAHAN

**ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
TOKOPEDIA MENGGUNAKAN MODEL BILSTM**

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi S-2 Teknologi Informasi

Oleh:

ERIK WARDI
NIM. 231231046

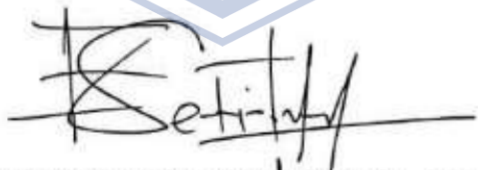
Disetujui Oleh:

Dosen Pembimbing


Dr. Ronsen Purba, M.Sc

Medan, 3 September 2025
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-2 Teknologi Informasi


Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-2 *) Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Erik Wardi

NIM : 231231046

Saya telah melaksanakan penelitian dan penulisan Tugas Akhir / Tesis *) dengan judul dan tempat penelitian sebagai berikut:

Judul Tugas Akhir / Tesis *) : Analisis Sentimen Ulasan Pengguna Aplikasi Tokopedia Menggunakan Model BiLSTM

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Tesis tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis 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 Tesis 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 Tesis 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 Tesis 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, 02 Juli 2025

Saya yang membuat pernyataan,



Erik Wardi

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI TOKOPEDIA MENGGUNAKAN MODEL BiLSTM

Abstrak

Perkembangan teknologi informasi telah mendorong peningkatan penggunaan aplikasi e-commerce, salah satunya Tokopedia sebagai salah satu platform terbesar di Indonesia. Ulasan pengguna pada platform seperti Google Play Store menyediakan informasi berharga tentang kepuasan dan pengalaman pengguna. Namun, tingginya volume data dan keragaman bahasa dalam ulasan menyebabkan tantangan dalam mengidentifikasi sentimen secara akurat. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Tokopedia berbahasa Indonesia ke dalam lima kategori: sangat positif, positif, netral, negatif, dan sangat negatif menggunakan pendekatan Bidirectional Long Short-Term Memory (BiLSTM). Dataset yang digunakan berjumlah 30.000 ulasan yang diperoleh melalui teknik web scraping. Proses analisis mencakup tahapan preprocessing teks (casefolding, cleaning, tokenization, stopword removal, stemming, dan normalization), pelabelan data menggunakan metode leksikal, serta transformasi teks ke bentuk vektor menggunakan word embedding berbasis IndoBERT. Data kemudian dibagi menjadi data latih dan data uji, dan dilakukan pelatihan model klasifikasi BiLSTM. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score melalui confusion matrix.

Kata kunci: *analisis sentiment, BiLSTM, IndoBERT, ulasan pengguna, aplikasi Tokopedia*

Abstract

The advancement of information technology has led to an increase in the use of e-commerce applications, with Tokopedia being one of the largest platforms in Indonesia. User reviews on platforms such as Google Play Store provide valuable insights into user satisfaction and experience. However, the high volume and linguistic diversity of these reviews present challenges in accurately identifying sentiment. This study aims to classify the sentiment of Indonesian-language user reviews of the Tokopedia application into five categories: very positive, positive, neutral, negative, and very negative using the Bidirectional Long Short-Term Memory (BiLSTM) approach. The dataset used comprises 30,000 reviews collected through web scraping techniques. The analysis process includes several text preprocessing stages (casefolding, cleaning, tokenization, stopword removal, stemming, and normalization), data labeling using a lexical method, and text transformation into vector representations using IndoBERT-based word embeddings. The data is then split into training and testing sets, followed by training a BiLSTM classification model. The model's performance is evaluated using the metrics of accuracy, precision, recall, and F1-score through a confusion matrix.

Keywords: *sentiment analysis, BiLSTM, IndoBERT, user reviews, Tokopedia application.*

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya, penulis dapat menyelesaikan p

1. Bapak Dr. Ronsen Purba, M.Sc selaku Dosen Pembimbing, yang telah membimbing penulis dalam proses penyelesaian tesis ini.
2. Ibu Sofiana Nurjanah, S.Kom., M.T.I., selaku Dosen Pendamping, yang juga telah membimbing penulis dalam proses penyelesaian tesis ini.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D., selaku Ketua Program Studi S-2 Teknologi Informasi Universitas Mikroskil Medan
6. Bapak atau Ibu Dosen Universitas Mikroskil Medan yang telah membantu proses penulisan tesis ini.
7. Anggota keluarga, teman, saudara dan semua pihak yang terus memberikan dukungan penuh kepada penulis selama proses penulisan tesis ini

Penulis menyadari bahwa masih terdapat berbagai kekurangan dalam penulisan tesis ini. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diterima. Akhir kata, penulis berharap tesis ini dapat memberikan manfaat bagi masyarakat. Terima kasih.

Medan, 02 Juli 2025

Penulis,

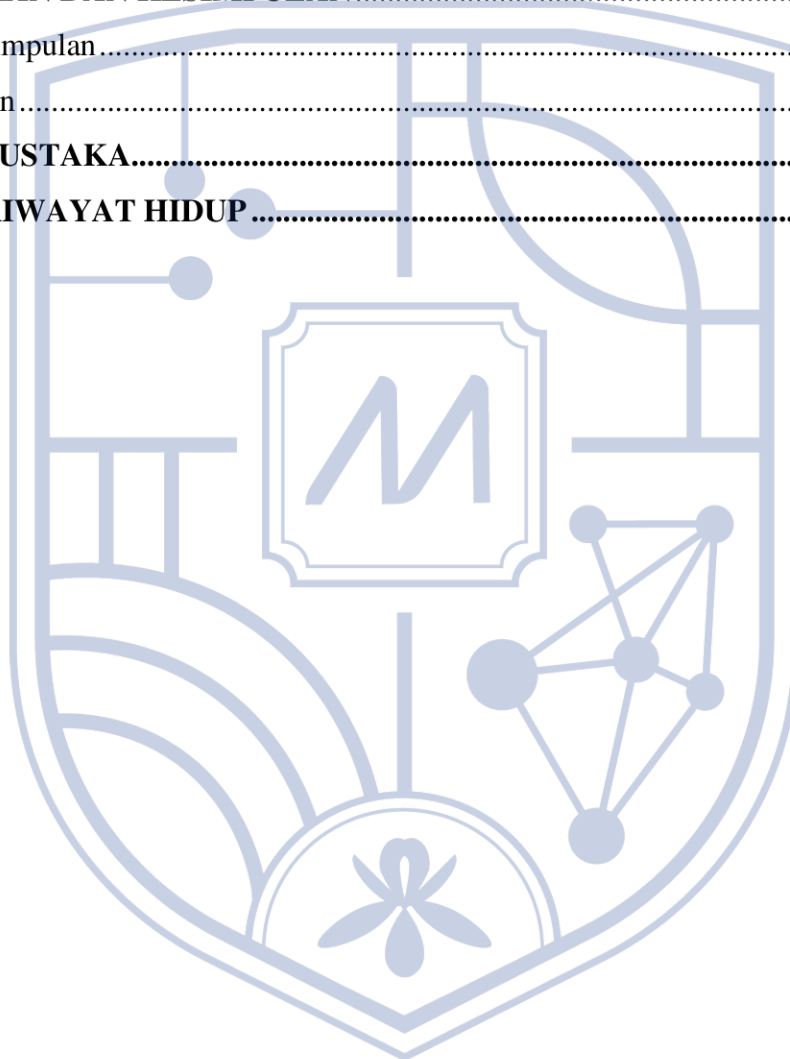


Erik Wardi

DAFTAR ISI

DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Manfaat.....	2
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR.....	4
2.1 Tinjauan Pustaka.....	4
2.1.1 Analisis Sentimen	4
2.1.2 <i>Natural Language Processing</i> (NLP).....	5
2.1.3 <i>Deep Learning</i>	5
2.1.4 <i>Bidirectional Encoder Representations from Transformers</i> (BERT)	6
2.1.5 <i>Fine-Tuning IndoBERT</i>	7
2.1.6 <i>Long Short-Term Memory</i> (LSTM)	8
2.1.5 <i>Bidirectional Long Short-Term Memory</i> (BiLSTM)	9
2.1.6 <i>Confussion Matrix</i>	10
2.2 Penelitian Terdahulu.....	11
2.3 Kerangka Pikir Pemecahan Masalah	13
BAB III METODOLOGI PENELITIAN	15
3.1 Analisis Masalah.....	15
3.2 Pengumpulan Data.....	16
3.3 Metode Penelitian	16
3.4 Alat-Alat Penelitian	19
3.5 Teknik Analisis Data	20
BAB IV HASIL DAN PEMBAHASAN	21
4.1 Hasil.....	21
4.1.1 <i>Dataset</i> yang digunakan	23

4.1.2 Data Pre-processing	20
4.1.3 Data Labeling	24
4.1.4 Word Embedding	27
4.1.5 Data Splitting	28
4.1.6 BiLSTM Classification.....	28
4.1.6 Training dan Evaluation	30
4.2 Pembahasan	31
BAB V SARAN DAN KESIMPULAN.....	36
5.1 Kesimpulan.....	36
5.2 Saran.....	36
DAFTAR PUSTAKA.....	38
DAFTAR RIWAYAT HIDUP	40



DAFTAR GAMBAR

Gambar 2.1 Diagram proses IndoBERT.....	7
Gambar 2.2 Diagram LSTM.....	8
Gambar 2.3 Diagram BiLSTM.....	9
Gambar 2.4 <i>Confusion Matrix</i>	10
Gambar 2.5 Kerangka konseptual	13
Gambar 3.1 <i>Flow chart</i> penelitian	17
Gambar 4.1 <i>Syntax Crawling</i> data ulasan.....	22
Gambar 4.2 <i>Casefolding and cleaning</i>	23
Gambar 4.3 <i>Tokenization</i>	24
Gambar 4.4 <i>Stopwords removal</i>	24
Gambar 4.5 <i>Text stemming</i>	24
Gambar 4.6 <i>Normalization</i>	24
Gambar 4.7 <i>Syntax labeling</i>	26
Gambar 4.8 Nilai persentase pelabelan pada data ulasan.....	27
Gambar 4.9 <i>Syntax word embedding</i> dengan IndoBERT.....	28
Gambar 4.10 <i>Syntax</i> untuk <i>data splitting</i>	28
Gambar 4.11 <i>Syntax</i> untuk persiapan pembelajaran dengan BiLSTM	29
Gambar 4.12 <i>Syntax</i> klasifikasi dengan BiLSTM.....	30
Gambar 4.13 <i>Syntax</i> untuk <i>training</i> dan <i>evaluation</i>	31
Gambar 4.14 Hasil Evaluasi <i>Confussion Matrix</i>	32
Gambar 4.15 Hasil Evaluasi <i>precision</i> , <i>recall</i> , dan <i>f1-score</i>	33
Gambar 4.16 Grafik pelatihan <i>epoch</i>	34

DAFTAR TABEL

Tabel 2.1 <i>Literature Review</i> Analisis Sentimen	11
Tabel 3.1 Contoh Data Ulasan	16
Tabel 3.2 Contoh <i>Casefolding</i>	17
Tabel 3.3 Contoh <i>Data Cleaning</i>	18
Tabel 3.4 Contoh <i>Stopword removal</i>	18
Tabel 3.5 Contoh <i>Stemming</i>	18
Tabel 4.1 Tabel <i>Dataset</i>	21
Tabel 4.2 Hasil <i>pre-processing</i> dari <i>dataset</i>	23
Tabel 4.3 Hasil <i>labeling</i>	25

