

**ANALISIS SENTIMEN MENGGUNAKAN MODEL
CONCATENATION CLASSIFICATION INDOBERT UNTUK
ULASAN PENGGUNA APLIKASI MYTELKOMSEL DI PLAYSTORE**

TESIS

Oleh:

**TRI FITRIA NINGSIH
NIM. 231231062**



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

**SENTIMENT ANALYSIS USING CONCATENATION
CLASSIFICATION INDOBERT MODEL FOR MYTELKOMSEL
APPLICATION REVIEW ON PLAYSTORE**

THESIS

By:

**TRI FITRIA NINGSIH
ID NUMBER. 231231062**



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

LEMBARAN PENGESAHAN

**ANALISIS SENTIMEN MENGGUNAKAN MODEL
CONCATENATION CLASSIFICATION INDOBERT UNTUK
ULASAN PENGGUNA APLIKASI MYTELKOMSEL DI PLAYSTORE**

TESIS

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

Oleh:

TRI FITRIA NINGSIH
NIM. 231231062

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 : Tri Fitria Ningsih
NIM : 231231062

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

Judul Tugas Tesis : Analisis Sentimen Menggunakan Model Concatenation Classification IndoBERT Untuk Ulasan Pengguna Aplikasi MyTelkomsel di Playstore

Tempat Penelitian : -
Alamat Tempat Penelitian : -
No. Telp. Tempat Penelitian : -

Selubungan 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, 1 Juli 2025

Saya yang membuat pernyataan,



Tri Fitria Ningsih

ANALISIS SENTIMEN MENGGUNAKAN MODEL CONCATENATION CLASSIFICATION INDOBERT UNTUK ULASAN PENGGUNA APLIKASI MYTELKOMSEL DI PLAYSTORE

Abstrak

Di era digital, ulasan pengguna aplikasi mobile menjadi sumber data penting untuk mengevaluasi dan meningkatkan kualitas layanan. Penelitian ini bertujuan menganalisis sentimen pengguna terhadap aplikasi MyTelkomsel di Google PlayStore menggunakan pendekatan Concatenation Classification IndoBERT, yaitu penggabungan embedding dari model IndoBERT dan FastText guna memperkaya representasi semantik teks. Sebanyak 30.000 ulasan berbahasa Indonesia dari tahun 2022–2024 dikumpulkan melalui web scraping dengan pustaka google-play-scraper. Data diproses melalui tahapan preprocessing (normalisasi, pembersihan, penghapusan stopword, dan stemming) serta pelabelan sentimen ke dalam lima kategori: sangat negatif, negatif, netral, positif, dan sangat positif berdasarkan skor polaritas. Pemodelan dilakukan dengan menggabungkan vektor token [CLS] dari IndoBERT (768 dimensi) dan FastText (300 dimensi), menghasilkan vektor 1068 dimensi. Dataset dibagi menjadi 80% data latih, 20% data uji dan 80% data latih, 10% data uji dan 10% data validasi. Evaluasi menggunakan metrik akurasi, precision, recall, dan F1-score menunjukkan performa tinggi dengan akurasi 95% pada 2022–2023 dan meningkat menjadi 96% pada 2024. Hasil ini menunjukkan bahwa pendekatan concatenation IndoBERT dapat meningkatkan akurasi klasifikasi sentimen secara signifikan dan efektif dalam menangani teks ulasan yang tidak terstruktur..

Kata kunci: Analisis sentimen, IndoBERT, Concatenation Classification, FastText, MyTelkomsel, NLP, Google PlayStore.

Abstract

In the digital era, user reviews of mobile applications have become an important data source for evaluating and improving service quality. This study aims to analyze user sentiment toward the MyTelkomsel application on Google PlayStore using the Concatenation Classification IndoBERT approach, which combines embeddings from IndoBERT and FastText to enrich the semantic representation of text. A total of 30,000 user reviews in Indonesian from 2022 to 2024 were collected through web scraping using the google-play-scraper library. The data was processed through several preprocessing stages, including normalization, cleaning, stopword removal, and stemming, followed by sentiment labeling into five categories: very negative, negative, neutral, positive, and very positive, based on polarity scores. Modeling was performed by combining the [CLS] token vector from IndoBERT (768 dimensions) and the FastText vector (300 dimensions), resulting in a 1068-dimensional vector. The dataset was split into 80% training data, 20% testing data and 80% training data, 10% testing data and 10% validation data . Evaluation using metrics such as accuracy, precision, recall, and F1-score showed high performance, with accuracy reaching 95% in 2022–2023 and increasing to 96% in 2024. These results indicate that the IndoBERT concatenation approach significantly improves sentiment classification accuracy and is effective in handling unstructured user review texts..

Keywords: Sentiment analysis, IndoBERT, Concatenation Classification, FastText, MyTelkomsel, NLP, Google PlayStore

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya, penulis dapat menyelesaikan tesis yang berjudul “**Analisis Sentimen Menggunakan Model Concatenation Classification Indobert Untuk Ulasan Pengguna Aplikasi MyTelkomsel di PlayStore**” dengan baik.

Tesis ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Magister Teknologi Informasi di Universitas Mikroskil Medan.

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam proses penyusunan tesis ini. Secara khusus, penulis mengucapkan terima kasih kepada :

1. Bapak Dr. Ronsen Purba, M.Sc., selaku Dosen Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
2. Bapak Muhammad Fermi Pasha, B.Sc., M.Sc., PhD., selaku Dosen Pendamping yang 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, MMSI., 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, 1 Juli 2025

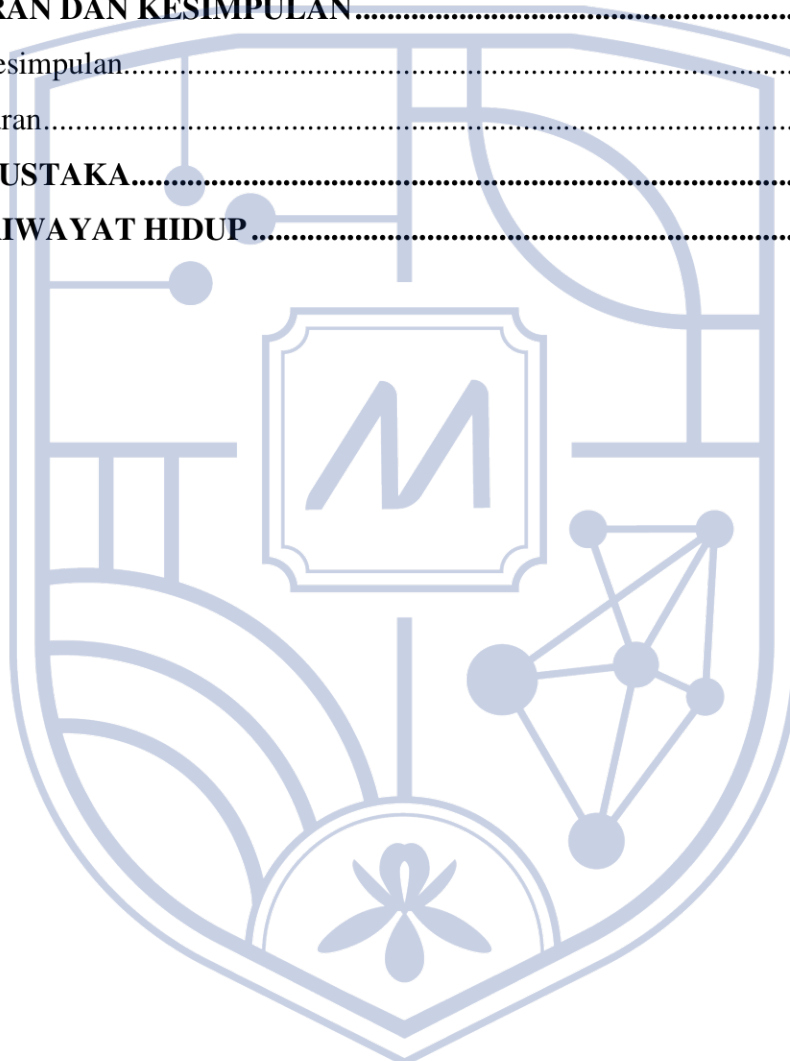
Penulis,

Tri Fitria Ningsih

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	iv
DAFTAR TABEL	iv
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>	5
2.1.3 BERT (<i>Bidirectional Encoder Representations from Transformers</i>)	6
2.1.4 IndoBERT	7
2.1.5 IndoBERT-BC.....	8
2.1.6 IndoBERT-BiC.....	8
2.1.7 IndoBERT-CC (<i>Concatenation Classification</i>)	9
2.1.8 <i>Confussion Matrix</i>	10
2.2 Penelitian Terdahulu.....	12
2.3 Kerangka Pikir Pemecahan Masalah	15
BAB III METODOLOGI PENELITIAN	17
3.1 Analisis Masalah	17
3.2 Pengumpulan Data	18
3.3 Metode Penelitian.....	19
3.4 Alat-Alat Penelitian	24
3.5 Teknik Analisis Data	24

BAB IV HASIL DAN PEMBAHASAN	26
4.1 Hasil	26
4.1.1 Dataset yang Digunakan.....	26
4.1.2 Menampilkan Hasil <i>Pre-Processing Data</i>	28
4.1.3 Menampilkan Hasil <i>Labeling Data</i>	29
4.1.4 Pemodelan <i>IndoBERT Concatenation Classification</i>	30
4.2 Pembahasan.....	36
BAB V SARAN DAN KESIMPULAN.....	40
5.1 Kesimpulan.....	40
5.2 Saran.....	41
DAFTAR PUSTAKA.....	42
DAFTAR RIWAYAT HIDUP	44



DAFTAR GAMBAR

Gambar 2.1 Arsitektur BERT	7
Gambar 2.2 Arsitektur Model IndoBERT-BC	8
Gambar 2.3 Arsitektur IndoBERT- BiC	9
Gambar 2.4 Arsitektur Model IndoBERT-CC	10
Gambar 2.5 <i>Confusion Matrix</i>	10
Gambar 2.6 Kerangka Konseptual	15
Gambar 3.1 Alur Metode Penelitian	19
Gambar 4.1 <i>Syntax Crawling Data Ulasan</i>	27
Gambar 4.2 Dataset yang Telah Melalui <i>Pre-Processing</i>	28
Gambar 4.3 <i>Syntax Pre-Processing Data</i>	28
Gambar 4.4 Hasil <i>Labeling Data</i>	29
Gambar 4.5 <i>Syntax Labeling Data</i>	29
Gambar 4.6 Hasil Distribusi Labeling Data Ulasan	30
Gambar 4.7 <i>Syntax Pemisahan Data</i>	31
Gambar 4.8 <i>Load Model</i>	31
Gambar 4.9 <i>Embedding Function</i>	32
Gambar 4.10 Proses <i>Concatenation</i>	33
Gambar 4.11 <i>Syntax Pelatihan dan Evaluasi</i>	34
Gambar 4.12 Hasil Evaluasi <i>Confussion Matrix</i> Ulasan Tahun 2022	34
Gambar 4.13 Hasil Evaluasi <i>Confussion Matrix</i> Ulasan Tahun 2023	35
Gambar 4.14 Hasil Evaluasi <i>Confussion Matrix</i> Ulasan Tahun 2024	35
Gambar 4.15 Grafik Perbandingan Akurasi Berdasarkan <i>Learning Rate</i> Pengujian 1.....	37
Gambar 4.15 Grafik Perbandingan Akurasi Berdasarkan <i>Learning Rate</i> Pengujian 2.....	38

DAFTAR TABEL

Tabel 2.1 Beberapa Penelitian Analisis Sentimen	13
Tabel 3.1 Contoh Data Ulasan.....	18
Tabel 3.2 Contoh <i>Case Folding</i>	20
Tabel 3.3 Contoh <i>Data Cleaning</i>	20
Tabel 3.4 Contoh <i>Stopwords</i>	20
Tabel 3.5 Contoh <i>Stemming</i>	21
Tabel 3.6 Contoh <i>Normalization</i>	21
Tabel 3.7 Tabel Nilai Polaritas.....	22
Tabel 3.8 Contoh Kamus <i>Lexicon</i>	22
Tabel 3.9 Contoh <i>Tokenization</i>	23
Tabel 4.1 Tabel Dataset.....	26
Tabel 4.2 Tabel Hasil Pengujian 1.....	36
Tabel 4.3 Tabel Hasil Pengujian 2.....	38