

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE FIT MENGGUNAKAN BERT DAN BERTOPIC

SKRIPSI

Oleh:

KENRICK FYLAN

NIM. 221110113

JONATHAN

NIM. 221110261

IRFANDI

NIM. 221110290



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

SENTIMENT ANALYSIS OF GOOGLE FIT USER REVIEWS USING BERT AND BERTOPIC

UNDERGRADUATE THESIS

By:

**KENRICK FYLAN
ID NUMBER. 221110113
JONATHAN
ID NUMBER. 221110261
IRFANDI
ID NUMBER. 221110290**



**UNDERGRADUATE PROGRAM OF COMPUTER SCIENCE
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE
FIT MENGGUNAKAN BERT DAN BERTOPIC

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknik Informatika

Oleh:

KENRICK FYLAN
NIM. 221110113
JONATHAN
NIM. 221110261
IRFANDI
NIM. 221110290

Disetujui Oleh:

Dosen Pembimbing I,


Sumaryo Winardi, S.Kom., M.T.

Dosen Pembimbing II,


Sio Jurnalis Pipin, S.Kom., M.Kom.

Medan, 14 Juli 2026
Diketahui dan Disahkan Oleh:


UNIKMA Program Studi
S-1 Teknik Informatika
MIKROSKIL

Charles Juliandy, S.Kom., M.Kom

HALAMAN PERNYATAAN

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

Nama : Kenrick Fylan

NIM : 221110113

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

Judul Skripsi : ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
GOOGLE FIT MENGGUNAKAN BERT DAN BERTOPIC

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi 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 Skripsi 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 Skripsi 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 Skripsi 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, 30 Juni 2026

Saya yang membuat pernyataan,



Kenrick Fylan

HALAMAN PERNYATAAN

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

Nama : Jonathan

NIM : 221110261

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

Judul Skripsi : Analisis Sentimen Ulasan Pengguna Aplikasi Google Fit Menggunakan BERT dan BERTopic

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi 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 Skripsi 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 Skripsi 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 Skripsi 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, 30 Juni 2026

Saya yang membuat pernyataan,



Jonathan

HALAMAN PERNYATAAN

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

Nama : Irfandi

NIM : 221110290

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

Judul Skripsi : ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI
GOOGLE FIT MENGGUNAKAN BERT DAN BERTOPIC

Tempat Penelitian : -

Alamat Tempat Penelitian : -

No. Telp. Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi 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 Skripsi 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 Skripsi 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 Skripsi 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, 30 Juni 2026

Saya yang membuat pernyataan,


Irfandi

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE FIT MENGGUNAKAN BERT DAN BERTOPIC

Abstrak

Perkembangan aplikasi kebugaran digital seperti Google Fit telah menghasilkan jumlah ulasan pengguna yang bervariasi dan mengandung sentimen beragam. Analisis terhadap ulasan tersebut penting untuk memahami tingkat kepuasan pengguna serta mengidentifikasi aspek aplikasi yang memerlukan perbaikan. Penelitian ini bertujuan mengembangkan sistem berbasis web yang mampu melakukan analisis sentimen dan pemodelan topik terhadap ulasan pengguna Google Fit menggunakan metode Bidirectional Encoder Representations from Transformers (BERT) dan BERTopic. Data penelitian berupa 10.928 ulasan berbahasa Inggris yang diperoleh melalui proses web scraping dari Google Play Store. Data tersebut kemudian dibagi menjadi data latih sebesar 70%, data uji sebesar 20%, dan data validasi sebesar 10%. Proses penelitian meliputi pengumpulan data, pelabelan sentimen menggunakan VADER, preprocessing, pelatihan model BERT untuk klasifikasi sentimen, serta pemodelan topik menggunakan BERTopic. Kinerja model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score, sedangkan kualitas topik dievaluasi menggunakan coherence score. Hasil pengujian model BERT pada 2.080 data uji menghasilkan performa klasifikasi dengan akurasi sebesar 90.67%, presisi sebesar 90.45%, recall sebesar 90.51%, dan F1-score sebesar 90.45%. Sementara itu hasil pemodelan topik BERTopic berhasil menghasilkan 16 topik utama dengan nilai koherensi rata-rata sebesar 0.6586.

Kata kunci: Analisis Sentimen, BERT, BERTopic, Natural Language Processing, Google Fit

Abstract

The growth of digital fitness apps such as Google Fit has resulted in a wide variety of user reviews expressing diverse sentiments. Analyzing these reviews is essential for understanding user satisfaction levels and identifying aspects of the app that require improvement. This study aims to develop a web-based system capable of performing sentiment analysis and topic modeling on Google Fit user reviews using the Bidirectional Encoder Representations from Transformers (BERT) and BERTopic methods. The research data consists of 10,928 English-language reviews obtained through web scraping from the Google Play Store. The data were split into 70% training data, 20% testing data, and 10% validation data. The research process included data collection, sentiment labeling using VADER, preprocessing, training the BERT model for sentiment classification, and topic modeling using BERTopic. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, while topic quality was evaluated using a coherence score. Testing the BERT model on 2,080 test data points yielded classification performance with an accuracy of 90.67%, precision of 90.45%, recall of 90.51%, and an F1-score of 90.45%. Meanwhile, the BERTopic topic modeling successfully identified 16 main topics with a mean coherence score of 0.6586.

Keywords: Sentiment Analysis, BERT, BERTopic, Natural Language Processing, Google Fit

KATA PENGANTAR

Puji dan syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul "Analisis Sentimen Ulasan Pengguna Aplikasi Google Fit Menggunakan BERT dan BERTopic".

Tugas Akhir ini disusun untuk melengkapi persyaratan meraih gelar Sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.

Penulis juga ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dosen Pembimbing I dan Dekan Fakultas Informatika Universitas Mikroskil Medan.
2. Bapak Sio Jurnal Pipin, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
5. Seluruh dosen dan staf pengajar yang telah membimbing kami selama menempuh perkuliahan di Universitas Mikroskil Medan.
6. Orang tua dan keluarga penulis yang telah memberikan dukungan dalam bentuk moral maupun materil.
7. Semua teman dan pihak yang tidak dapat disebutkan satu per satu oleh penulis yang telah memberikan bantuan dan dukungan dalam menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna dikarenakan keterbatasan pengalaman dan pengetahuan penulis. Oleh karena itu, penulis sangat menghargai kritik dan saran yang diberikan oleh pihak lainnya. Tugas Akhir ini juga diharapkan dapat memberikan wawasan dan manfaat positif bagi pihak-pihak yang membutuhkan. Terima kasih.

Medan, 14 Juli 2026

Penulis,

Kenrick Fylan

Jonathan

Irfandi

DAFTAR ISI

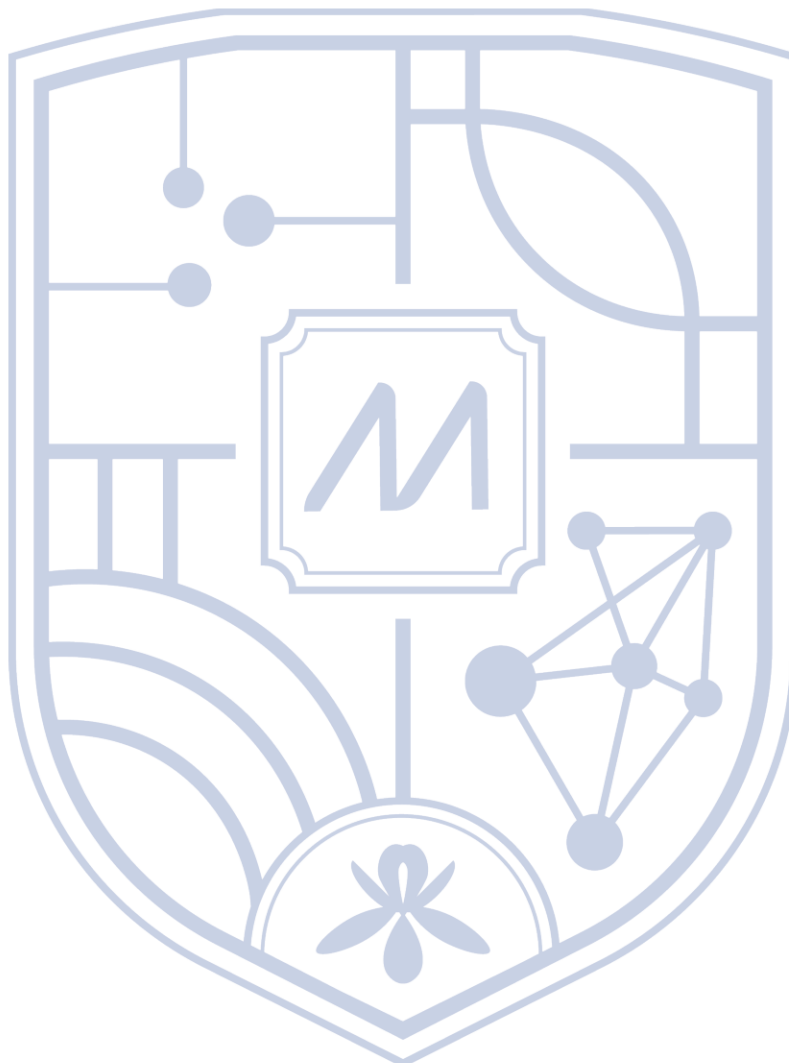
ABSTRAK.....	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Manfaat.....	3
1.5 Ruang Lingkup.....	3
BAB II KAJIAN LITERATUR.....	4
2.1 <i>Natural Language Processing</i> (NLP)	4
2.1.1 <i>Natural Language Understanding</i> (NLU).....	5
2.1.2 <i>Natural Language Generation</i> (NLG).....	7
2.2 Analisis Sentimen	8
2.3 <i>Machine Learning</i>	10
2.4 <i>Web Scraping</i>	11
2.5 <i>Data Labeling</i>	13
2.6 <i>Valence Aware Dictionary and Sentiment Reasoner</i> (VADER)	15
2.7 <i>Bidirectional Encoder Representations from Transformers</i> (BERT).....	16
2.7.1 Representasi Input	18
2.7.2 <i>Pre-training</i>	20
2.7.3 <i>Fine-tuning</i>	21
2.8 <i>Topic Modeling</i>	22
2.9 BERTopic	23
2.10 Evaluasi Model	29
2.11 Google Fit	30
2.12 Penelitian Terdahulu.....	32
BAB III TAHAPAN PELAKSANAAN	35
3.1 Kerangka Tahapan Pelaksanaan.....	35
3.2 Pengumpulan Data	36
3.3 Analisis Proses	36

3.3.1 Pelabelan Data	37
3.3.2 <i>Preprocessing Data</i>	44
3.3.3 Pembagian Data	50
3.3.4 Ekstraksi Embedding BERT	50
3.3.5 Klasifikasi Sentimen BERT	58
3.3.6 Pemodelan Topik Menggunakan BERTopic	62
3.4 Analisis Kebutuhan	70
3.4.1 Analisis Kebutuhan Fungsional	70
3.4.2 Analisis Kebutuhan Non-Fungsional	80
3.5 Perancangan	81
BAB IV HASIL DAN PEMBAHASAN	88
4.1 Hasil	88
4.1.1 Tampilan Halaman Utama Upload Dataset (Pengumpulan Data)	88
4.1.2 Tampilan Halaman Model Pelabelan Data	89
4.1.3 Tampilan Halaman Preprocessing BERT	90
4.1.4 Tampilan Halaman Evaluasi Model Klasifikasi Sentimen BERT	91
4.1.5 Tampilan Halaman Preprocessing BERTopic	93
4.1.6 Tampilan Halaman Evaluasi BERTopic (Integrasi Google Colab)	93
4.1.7 Tampilan Modul Hasil Analisis & Visualisasi	94
4.2 Hasil Pengujian Sistem	96
4.2.1 Hasil Pengujian Hyperparameter BERT	96
4.2.2 Hasil Evaluasi Model BERT pada Data Uji	103
4.2.3 Hasil Pemodelan Topik BERTopic	105
4.2.4 Hasil Sentimen Berdasarkan Topik	107
4.3 Rangkuman Hasil dan Pembahasan	109
BAB V PENUTUP	110
5.1 Kesimpulan	110
5.2 Saran	110
DAFTAR PUSTAKA	111
DAFTAR RIWAYAT HIDUP	117

DAFTAR GAMBAR

Gambar 2.1 Pohon Turunan dan Tingkatan Struktur Kalimat [32].....	6
Gambar 2.2 Fungsi <i>reviews</i> [51].....	12
Gambar 2.3 Contoh Hasil Skor Kalimat [58].....	16
Gambar 2.4 Arsitektur <i>Transformer</i> [60]	17
Gambar 2.5 Prosedur BERT dalam <i>Pre-Training</i> dan <i>Fine-Tuning</i> [61].....	18
Gambar 2.6 Representasi <i>Input</i> BERT [61].....	18
Gambar 2.7 Diagram Proses Topik Model [70]	23
Gambar 2.8 Alur kerja BERTopic [72]	24
Gambar 2.9 Tampilan Beranda Google Fit	31
Gambar 3.1 Alur Tahapan Penelitian	37
Gambar 3.2 Tahap Pelabelan Data	38
Gambar 3.3 Preprocessing untuk BERT(a) dan BERTopic(b)	45
Gambar 3.4 Tahapan Tokenisasi Teks oleh BERT	51
Gambar 3.5 Rancangan Use Case Diagram.....	71
Gambar 3.6 Rancangan Halaman Tampilan Input Data.....	82
Gambar 3.7 Rancangan Tampilan Pelabelan Data	83
Gambar 3.8 Rancangan Tampilan Preprocessing BERT.....	84
Gambar 3.9 Rancangan Tampilan Evaluasi BERT	84
Gambar 3.10 Rancangan Tampilan Preprocessing BERTopic.....	85
Gambar 3.11 Rancangan Tampilan Evaluasi BERTopic	86
Gambar 3.12 Rancangan Tampilan Hasil Analisis dan Visualisasi	87
Gambar 4.1 Tampilan Proses Pemilihan Berkas CSV melalui File Explorer	88
Gambar 4.2 Tampilan Menu Utama Upload Dataset dengan Berkas Terpilih.....	89
Gambar 4.3 Tampilan Pratinjau Dataset yang Berhasil Dimuat	89
Gambar 4.4 Tampilan Halaman Menu Pelabelan Data Utama	90
Gambar 4.5 Tampilan Halaman Preprocessing BERT Sebelum Eksekusi	90
Gambar 4.6 Tampilan Log Hasil Transformasi Teks Preprocessing BERT	91
Gambar 4.7 Tampilan Laporan Klasifikasi (Classification Report) Model BERT	91
Gambar 4.8 Tampilan Antarmuka Confusion Matrix Hasil Evaluasi BERT	92
Gambar 4.9 Tampilan Antarmuka Tabel Detail Prediksi Sentimen (Predictions Detail) ...	92
Gambar 4.10 Tampilan Hasil Modul Preprocessing BERTopic	93
Gambar 4.11 Tampilan Menu Fetch Hasil BERTopic Modeling.....	94
Gambar 4.12 Tampilan Grafik Distribusi Dokumen dan Tabel Nilai Koherensi Topik	94

Gambar 4.13 Tampilan Dashboard Grafik Sebaran Sentimen pada Setiap Klaster Topik..	95
Gambar 4.14 Tampilan Diagram Lingkaran Distribusi Sentimen Global dan Tabel Ringkasan	95
Gambar 4.15 Tampilan Fitur Analisis Detail dan Sampel Teks Ulasan per-Topik.....	96



DAFTAR TABEL

Tabel 2.1 Evaluasi Confusion Matrix.....	29
Tabel 2.2 Penelitian Terdahulu.....	32
Tabel 3.1 Sample Dataset	36
Tabel 3.2 Preprocessing VADER.....	39
Tabel 3.3 Tokenisasi VADER.....	39
Tabel 3.4 Scoring menggunakan kamus Lexicon.....	40
Tabel 3.5 Penerapan Heuristik VADER.....	41
Tabel 3.6 Perhitungan Compound Score.....	42
Tabel 3.7 Pelabelan Data.....	44
Tabel 3.8 Cleansing Data.....	45
Tabel 3.9 Case Folding.....	46
Tabel 3.10 Tahapan Validasi Data.....	47
Tabel 3.11 Tokenisasi.....	47
Tabel 3.12 Stopword Removal.....	48
Tabel 3.13 Lemmatization.....	49
Tabel 3.14 Perbandingan Hasil Preprocessing data BERT dan BERTopic	49
Tabel 3.15 Pembagian Data.....	50
Tabel 3.16 Pembentukan Token Embedding.....	51
Tabel 3.17 Pembentukan Positional Embedding.....	53
Tabel 3.18 Pembentukan Segment Embedding.....	54
Tabel 3.19 Output Hasil Encoder BERT.....	56
Tabel 3.20 Output Logits.....	59
Tabel 3.21 Contoh Perhitungan <i>Term Frequency per Cluster</i> dan <i>Total Frequency Global</i>	67
Tabel 3.22 Contoh Perhitungan Nilai Logaritma.....	68
Tabel 3.23 Contoh Perhitungan c-TF-IDF untuk Topik 0.....	69
Tabel 3.24 Contoh Perhitungan c-TF-IDF untuk Topik 1.....	69
Tabel 3.25 Contoh Perhitungan c-TF-IDF untuk Topik 2.....	69
Tabel 3.26 Penjelasan Use Case Diagram	71
Tabel 3.27 Use Case Mengunggah Data	72
Tabel 3.28 Use Case Melabeli Data	73
Tabel 3.29 Use Case Melakukan Preprocessing BERT	74
Tabel 3.30 Use Case Melakukan Evaluasi BERT	75

Tabel 3.31 Use Case Melakukan Preprocessing BERTopic	77
Tabel 3.32 <i>Use Case</i> Melakukan Evaluasi BERTopic	78
Tabel 3.33 Use Case Melihat Hasil Analisis dan Visualisasi.....	80
Tabel 3.34 Analisis Kebutuhan Non-Fungsional PIECES	80
Tabel 4.1 Hasil Evaluasi Full Dataset Training Pada Data Pelatihan dan Data Validasi....	97
Tabel 4.2 Confusion Matrix Hasil Evaluasi Prediksi Sentimen	103
Tabel 4.3 Karakteristik Kluster Topik Utama BERTopic	105
Tabel 4.4 Contoh Data Ulasan Pengguna Berdasarkan Kluster Topik.....	108

