

ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TIKTOK MENGGUNAKAN METODE INDOBERT

SKRIPSI

Oleh:

SITI JULAIKA BR TARIGAN

NIM. 221130725

HOSEA SIREGAR

NIM. 221130821



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

SENTIMENT ANALYSIS OF CORETAX USER REVIEW ON TIKTOK USING THE INDOBERT METHOD

UNDERGRADUATE THESIS

By:

SITI JULAIKA BR TARIGAN

ID NUMBER. 221130725

HOSEA SIREGAR

ID NUMBER. 221130821



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

LEMBARAN PENGESAHAN

ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TIKTOK MENGGUNAKAN METODE INDOBERT

SKRIPSI

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

Oleh:

SITI JULAIKA BR TARIGAN
NIM. 221130725
HOSEA SIREGAR
NIM. 221130821

Disetujui Oleh:

Dosen Pembimbing I



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

Dosen Pembimbing II



Mustika Ulina, S.Kom., M.Kom.

Medan, 1 Agustus 2026
Diketahui dan Disahkan Oleh:



Ketua Program Studi
S1 Teknologi Informasi

Mustika 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 : Hosea Siregar

NIM : 221130821

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

Judul Skripsi : ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TIKTOK MENGGUNAKAN METODE INDOBERT

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, 1 Juni 2026

Saya yang membuat pernyataan,


METERAI TEMPEL
22CAOX066353645
Hosea Siregar

HALAMAN PERNYATAAN

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

Nama : Siti Julaika Br Tarigan
NIM : 221130725

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

Judul Skripsi : ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TIKTOK MENGGUNAKAN METODE INDOBERT

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, 1 Juni 2026

Saya yang membuat pernyataan,



Siti Julaika Br Tarigan

ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TIKTOK MENGGUNAKAN METODE INDOBERT

Abstrak

Sistem Coretax merupakan sistem administrasi perpajakan baru yang diluncurkan oleh Direktorat Jenderal Pajak (DJP) dan menuai beragam respons dari masyarakat, khususnya di platform media sosial TikTok. Penelitian ini bertujuan untuk menganalisis kecenderungan sentimen masyarakat terhadap implementasi sistem Coretax berdasarkan komentar pengguna TikTok menggunakan model IndoBERT. Data dikumpulkan melalui proses crawling dan melewati tahapan preprocessing meliputi cleaning, case folding, dan normalisasi. Pelabelan data dilakukan oleh dua anotator dengan hasil uji Cohen's Kappa sebesar $\kappa = 0,9889$ yang menunjukkan tingkat reliabilitas sangat tinggi. Model IndoBERT-base-p2 dilatih menggunakan teknik fine-tuning dengan penerapan class weight untuk menangani ketidakseimbangan kelas. Dari 5.518 komentar berlabel, sebanyak 78,3% bersentimen negatif dan 21,7% bersentimen positif. Model terbaik dengan konfigurasi Learning Rate = $3e-05$, Batch Size = 16, Epoch = 7, dan Dropout = 0,1 berhasil mencapai akurasi sebesar 96,92%, F1-score macro average 95,43%, dan recall kelas positif 92,00%. Hasil penelitian membuktikan bahwa IndoBERT dengan penerapan class weight merupakan pendekatan yang efektif untuk analisis sentimen opini publik terhadap layanan digital pemerintah berbasis data media sosial.

Kata kunci: analisis sentimen, Coretax, IndoBERT, TikTok, media sosial

Abstract

The Coretax system is a new tax administration system launched by the Directorate General of Taxes (DGT) that has received various responses from the public, particularly on the TikTok social media platform. This study aims to analyze public sentiment tendencies toward the implementation of the Coretax system based on TikTok user comments using the IndoBERT model. Data were collected through a crawling process and underwent preprocessing stages including cleaning, case folding, and normalization. Data labeling was performed by two annotators, yielding a Cohen's Kappa value of $\kappa = 0.9889$, indicating a very high level of reliability. The IndoBERT-base-p2 model was trained using fine-tuning techniques with class weight implementation to handle class imbalance. Of 5,518 labeled comments, 78.3% were negative and 21.7% were positive in sentiment. The best model configuration with Learning Rate = $3e-05$, Batch Size = 16, Epoch = 7, and Dropout = 0.1 achieved an overall accuracy of 96.92%, a macro average F1-score of 95.43%, and a positive class recall of 92.00%. The results demonstrate that IndoBERT with class weight implementation is an effective approach for public sentiment analysis toward government digital services based on social media data.

Keywords: sentiment analysis, Coretax, IndoBERT, TikTok, social media

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala kasih karunia, berkat, serta penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Sentimen Komentar TikTok terhadap Implementasi Coretax Menggunakan Model IndoBERT" dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Teknologi Informasi, Fakultas Informatika, Universitas Mikroskil Medan.

Penyusunan skripsi ini dilakukan karena penulis ingin mengkaji opini masyarakat terhadap implementasi sistem Coretax melalui analisis sentimen pada komentar pengguna TikTok. Selain itu, penelitian ini diharapkan dapat memberikan kontribusi dalam penerapan teknologi Natural Language Processing (NLP), khususnya menggunakan model IndoBERT untuk klasifikasi sentimen berbahasa Indonesia.

Selama proses penyusunan skripsi ini, penulis memperoleh banyak bimbingan, dukungan, bantuan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Wulan Sri Lestari, S.Kom., M.Kom., selaku Dosen Pembimbing I yang telah meluangkan waktu, memberikan arahan, masukan, serta bimbingan dengan penuh kesabaran selama proses penyusunan skripsi ini.
2. Ibu Mustika Ulina, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan bimbingan, saran, motivasi, dan perhatian sehingga penulis dapat menyelesaikan skripsi ini dengan baik, dan selaku Ketua Program Studi S-1 Teknologi Informasi Fakultas Informatika Universitas Mikroskil Medan.
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. Seluruh dosen dan staf Fakultas Informatika Universitas Mikroskil Medan yang telah memberikan ilmu, pengalaman, serta pelayanan yang baik selama penulis menempuh pendidikan.
6. Kedua orang tua, keluarga, dan seluruh kerabat yang senantiasa memberikan doa, kasih sayang, semangat, dukungan moral maupun material kepada penulis selama masa perkuliahan hingga penyusunan skripsi ini selesai.

7. Teman-teman seperjuangan Program Studi S-1 Teknologi Informasi Universitas Mikroskil yang telah memberikan dukungan, motivasi, dan bantuan selama proses penyusunan skripsi.
8. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan dan dukungan sehingga skripsi ini dapat diselesaikan.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak sebagai bahan perbaikan di masa yang akan datang. Penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan, khususnya di bidang analisis sentimen dan Natural Language Processing, maupun bagi pembaca dan pihak-pihak yang membutuhkan.

Medan, 1 Juli 2026

Penulis,

Hosea Siregar

Siti Julaika Br Tarigan

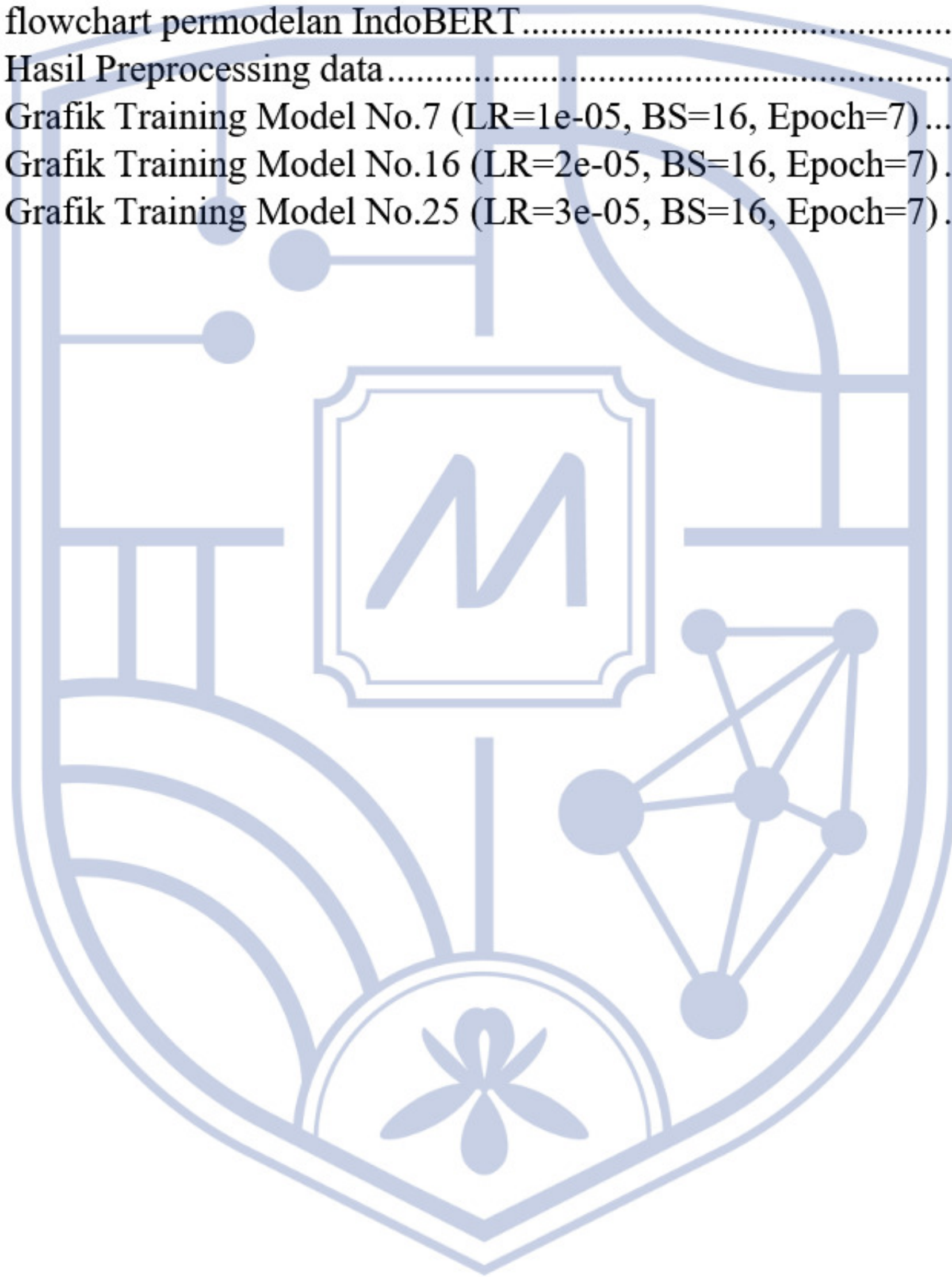
DAFTAR ISI

Abstrak	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR	5
2.1 Sistem Coretax	5
2.2 Media Sosial TikTok	6
2.3 Karakteristik Data Teks Media Sosial	8
2.4 Crawling Data	8
2.5 Natural Language Processing	10
2.6 Analisis Sentimen	11
2.7 Preprocessing	12
2.8 Pelabelan Data	13
2.9 Pembagian Data	15
2.10 BERT (Bidirectional Encoder Representations From Transformers)	17
2.11 IndoBERT (Indonesian <i>Bidirectional Encoder Representation From Transformers</i>)	21
2.12 Confusion Matrix	24
2.12.1 Precision	25
2.12.2 Recall	26
2.12.3 F1-Score	26
2.12.4 Accuracy	27
2.13 Word Cloud	27
BAB III TAHAPAN PELAKSANAAN	29
3.1 Pengumpulan Data	30
3.2 Analisis Proses	30

3.2.1 Preprocessing Data.....	31
3.2.2 Pelabelan Data	35
3.2.3 Pembagian Data	37
3.2.4 Pemodelan IndoBERT	39
3.2.5 Evaluasi Model	51
3.2.6 Word Cloud.....	53
BAB IV HASIL DAN PEMBAHASAN	54
4.1 Hasil	54
4.1.1 Hasil Pengumpulan Data.....	54
4.1.2 Hasil <i>Preprocessing</i> Data	55
4.1.3 Hasil Pelabelan Data dan Uji <i>Cohen's Kappa</i>	56
4.1.4 Hasil Pembagian Dataset	57
4.1.5 Hasil Penerapan <i>Class Weight</i>	58
4.1.6 Hasil Pembuatan Model.....	59
4.1.7 Hasil Evaluasi Model.....	65
4.2 Pembahasan.....	66
4.2.1 Distribusi Sentimen Komentar Pengguna TikTok terhadap Sistem Coretax.....	66
4.2.2 Kinerja Model IndoBERT dalam Analisis Sentimen.....	66
4.2.3 Analisis Topik Berdasarkan Frekuensi Kata pada Setiap Sentimen.....	67
BAB V PENUTUP	69
5.1 Kesimpulan	69
5.2 Saran	69
DAFTAR PUSTAKA	71
LAMPIRAN	77
DAFTAR RIWAYAT HIDUP	78

DAFTAR GAMBAR

Gambar 2. 1 Persentase Kepatuhan Wajib Pajak 2017 – 2023	5
Gambar 2. 2 Data Media Sosial yang Paling Banyak Diakses di Indonesia Tahun 2022	7
Gambar 2. 3 Arsitektur Model BERT	18
Gambar 2. 4 Ilustrasi tahapan Pre-training dan Fine-Tuning	19
Gambar 2. 5 Tahap Tokenisasi IndoBERT	22
Gambar 3. 1 Flowchart Analisis Proses.....	31
Gambar 3. 2 Flowchart Preprocessing Data	32
Gambar 3. 3 flowchart permodelan IndoBERT.....	39
Gambar 4. 1 Hasil Preprocessing data.....	55
Gambar 4. 2 Grafik Training Model No.7 (LR=1e-05, BS=16, Epoch=7).....	62
Gambar 4. 3 Grafik Training Model No.16 (LR=2e-05, BS=16, Epoch=7).....	62
Gambar 4. 4 Grafik Training Model No.25 (LR=3e-05, BS=16, Epoch=7).....	62



DAFTAR TABEL

Tabel 2. 1 Confusion Matrix untuk Klasifikasi Sentimen	25
Tabel 3. 1 Contoh Dataset yang digunakan.....	30
Tabel 3. 2 Contoh Proses Cleaning	33
Tabel 3. 3 Contoh Hasil Case Folding.....	34
Tabel 3. 4 Contoh Hasil Normalisasi.....	35
Tabel 3. 5 Hasil Pelabelan Sentimen.....	36
Tabel 3. 6 Hasil Tokenisasi, Konversi ke ID, dan Attention Mask (IndoBERT base-p2) ..	41
Tabel 4. 1 Hasil Pengumpulan Data	54
Tabel 4. 2 Distribusi Pelabelan dan Kesepakatan Antar Anotator	57
Tabel 4. 3 Hasil Pembagian Dataset.....	58
Tabel 4. 4 Tabel 4. 4 Nilai Class Weight untuk Setiap Kelas Sentimen	58
Tabel 4. 5 Eksperimen Pemilihan Learning Rate dan Batch Size	59
Tabel 4. 6 Perbandingan Tiga Model Terbaik Hasil Eksperimen	63
Tabel 4. 7 Konfigurasi Hyperparameter Fine-Tuning IndoBERT	64
Tabel 4. 8 Confusion Matrix Hasil Pengujian Model IndoBERT	65