

**OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS *STANCE
DETECTION* KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA**

SKRIPSI

Oleh:

**DONI CHARLES HUTAGAOL
NIM. 221112119
ADINDA DWI RAMADHANI
NIM. 221113708**



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

**INDOBERT OPTIMIZATION WITH SMOTE FOR STANCE DETECTION
ANALYSIS OF TIKTOK COMMENTS: A CASE STUDY OF HOGI
MINAYA**

FINAL RESEARCH

By:

**DONI CHARLES HUTAGAOL
ID. 221112119
ADINDA DWI RAMADHANI
ID. 221113708**



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

LEMBARAN PENGESAHAN
OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS *STANCE*
***DETECTION* KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA**

SKRIPSI

Dinjukan untuk Melengkapi Persyaratan Guna

Mendapatkan Gelar Sarjana

Program Studi S-1 Teknik Informatika

Oleh:

DONI CHARLES HUTAGAOL

NIM. 221112119

ADINDA DWI RAMADHANI

NIM. 221113708

Disetujui Oleh:

Dosen Pembimbing I,



Irpan Adiputra Pardosi, S.Kom., M.TI.

Dosen Pembimbing II,



Heru Kurniawan, S.Kom., M.Kom.

Medan, 20 Agustus 2026

Diketahui dan Disahkan Oleh:

Ketua Program Studi

S-1 Teknik Informatika,



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 : DONI CHARLES HUTAGAOL

NIM : 221112119

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

Judul Skripsi : OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS STANCE DETECTION KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA

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 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, 03 Agustus 2026

Saya yang membuat pernyataan,


Doni Charles Hutagaol

HALAMAN PERNYATAAN

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

Nama : ADINDA DWI RAMADHANI

NIM : 221113708

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

Judul Skripsi : OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS STANCE DETECTION KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA

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 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, 03 Agustus 2026

Saya yang membuat pernyataan,



Adinda Dwi Ramadhani

OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS STANCE DETECTION KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA

Abstrak

Kasus penetapan tersangka Hogi Minaya yang viral di TikTok memicu polarisasi opini publik, sehingga diperlukan pendekatan stance detection untuk mengidentifikasi keberpihakan komentar warganet, karena analisis sentimen konvensional rentan gagal mendeteksi sarkasme. Penelitian ini membangun model stance detection pada 3.970 komentar TikTok kasus Hogi Minaya menggunakan IndoBERT sebagai ekstraktor fitur kontekstual dan MLPClassifier sebagai model klasifikasi dua kelas: Kelas 0 (Support Hogi/Critical Police) dan Kelas 1 (Non-Support Hogi). Dataset mengalami ketidakseimbangan kelas dengan rasio $\pm 6,80:1$ (509 komentar Kelas 0, 3.461 komentar Kelas 1), sehingga diterapkan optimasi hyperparameter menggunakan GridSearchCV serta SMOTE pada ruang embedding IndoBERT. Pengujian pada 794 data uji dilakukan melalui tiga skenario: sebelum optimasi, setelah optimasi tanpa SMOTE, dan setelah penerapan SMOTE. Optimasi hyperparameter meningkatkan Accuracy dari 93,58% menjadi 93,83% dan Precision Kelas 0 dari 82,28% menjadi 89,55%, tetapi menurunkan Recall Kelas 0 dari 63,73% menjadi 58,82%. Setelah penerapan SMOTE, model menghasilkan Accuracy 93,07%, Precision Kelas 0 75,82%, Recall Kelas 0 67,65% (tertinggi di antara ketiga skenario), dan F1-Score Kelas 0 71,50%. Temuan ini menunjukkan optimasi hyperparameter saja belum mampu mengatasi ketidakseimbangan kelas, sedangkan SMOTE efektif meningkatkan kemampuan model mengenali kelas minoritas meskipun disertai penurunan Precision dan Accuracy. Model terbaik diimplementasikan dalam aplikasi web berbasis Streamlit sebagai sarana demonstrasi hasil penelitian.

Kata kunci: stance detection, IndoBERT, SMOTE, ketidakseimbangan kelas, TikTok

Abstract

The viral case of Hogi Minaya's suspect status on TikTok triggered public opinion polarization, requiring stance detection to identify comment alignment, as sentiment analysis often fails to detect sarcasm. This research builds a stance detection model on 3,970 TikTok comments on the Hogi Minaya case using IndoBERT as a contextual feature extractor and MLPClassifier as a two-class classifier: Class 0 (Support Hogi/Critical Police) and Class 1 (Non-Support Hogi). The dataset exhibited class imbalance with a ratio of $\pm 6.80:1$ (509 Class 0, 3,461 Class 1 comments), so hyperparameter optimization via GridSearchCV and SMOTE were applied to the IndoBERT embedding space. Testing on 794 samples covered three scenarios: before optimization, after optimization without SMOTE, and after SMOTE. Optimization increased Accuracy from 93.58% to 93.83% and Precision from 82.28% to 89.55%, but decreased Recall from 63.73% to 58.82% for Class 0. After SMOTE, the model achieved an Accuracy of 93.07%, Precision of 75.82%, Recall of 67.65% (the highest among the three scenarios), and an F1-Score of 71.50% for Class 0. These findings show hyperparameter optimization alone is insufficient for class imbalance, whereas SMOTE effectively improved minority-class recognition, albeit with lower Precision and Accuracy. The best model was implemented in a Streamlit-based web application to demonstrate the research results.

Keywords: stance detection, IndoBERT, SMOTE, class imbalance, TikTok

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa atas berkat, rahmat, dan penyertaan-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik yang berjudul "OPTIMASI INDOBERT DENGAN SMOTE UNTUK ANALISIS STANCE DETECTION KOMENTAR TIKTOK: STUDI KASUS HOGI MINAYA".

Tugas akhir ini bertujuan mengembangkan sistem analisis opini melalui pendekatan *stance detection* menggunakan IndoBERT dan SMOTE untuk mengatasi ketidakseimbangan data komentar TikTok pada kasus Hogi Minaya. Penelitian ini disusun sebagai syarat kelulusan Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan.

Penulis menyadari bahwa terselesaikannya skripsi ini tidak terlepas dari bimbingan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Irpan Adiputra Pardosi, S.Kom., M.TI., selaku Dosen Pembimbing I.
2. Bapak Heru Kurniawan, S.Kom., M.Kom., selaku Dosen Pembimbing II.
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 Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
6. Kedua orang tua dan keluarga tercinta, atas doa, dukungan moral, material, serta kasih sayang yang tak terhingga sepanjang masa perkuliahan hingga selesainya skripsi ini.
7. Teman-teman seperjuangan Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan, atas kebersamaan, motivasi, dan dukungannya selama ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Akhir kata, semoga karya ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan pembaca sekalian.

Medan, 03 Agustus 2026
Penulis,

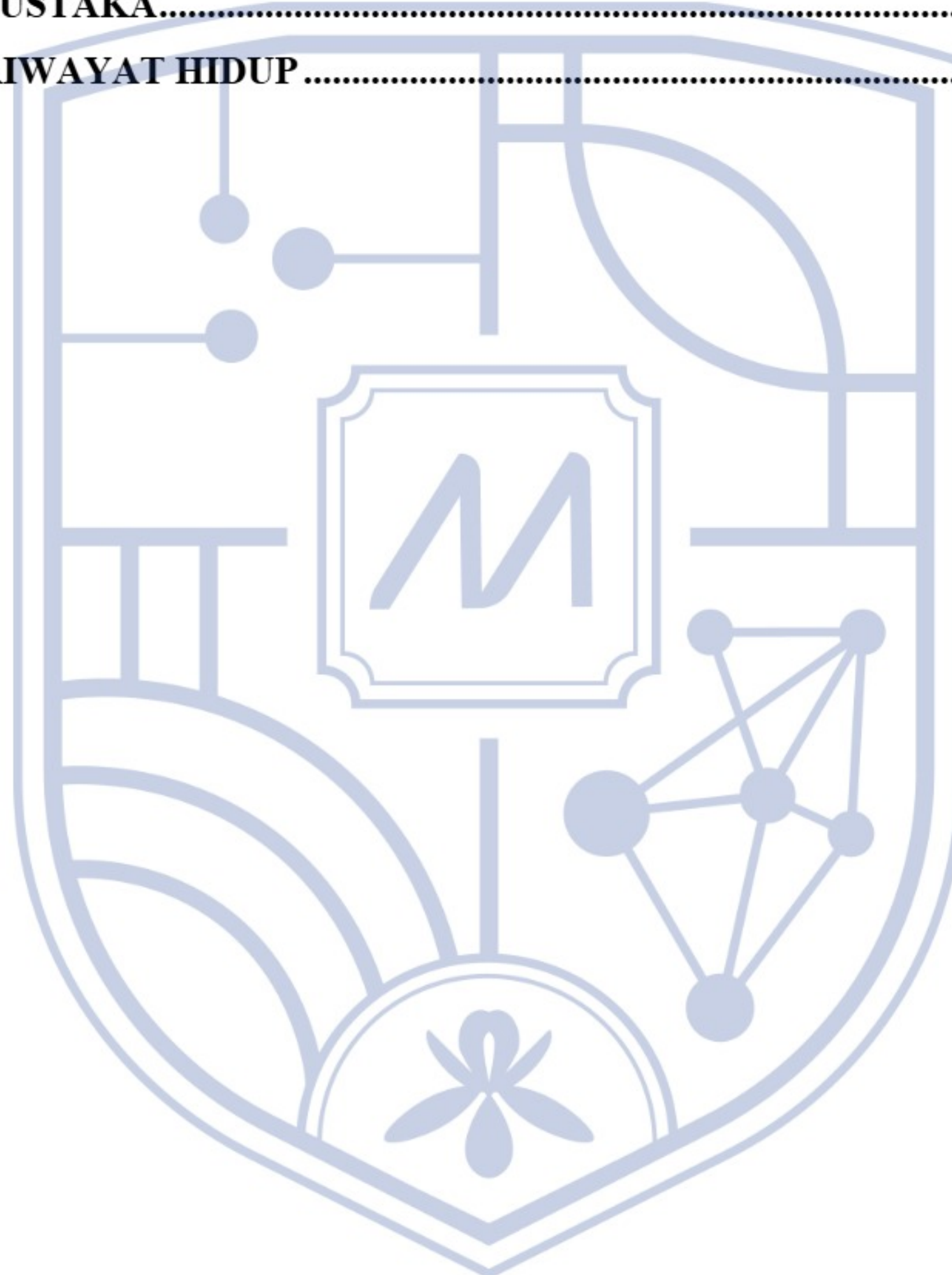
Doni Charles Hutagaol
Adinda Dwi Ramadhani

DAFTAR ISI

ABSTRAK.....	I
KATA PENGANTAR.....	II
DAFTAR ISI	III
DAFTAR GAMBAR.....	VI
DAFTAR TABEL	VII
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan	4
1.4 Manfaat	4
1.5 Ruang Lingkup.....	4
BAB II KAJIAN LITERATUR.....	6
2.1 Penelitian Terdahulu	6
2.2 Natural Language Processing (NLP)	9
2.3 Stance Detection dan Sentiment Analysis	13
2.4 Pemanfaatan Large Language Model sebagai Alat Bantu Anotasi.....	15
2.5 Arsitektur Transformer dan Mekanisme Attention.....	16
2.6 Model BERT.....	18
2.6.1 Representasi Input pada BERT	19
2.6.2 Model IndoBERT.....	20
2.7 Class Imbalance dan Algoritma SMOTE	22
2.7.1 Class Imbalance	23
2.7.2 Synthetic Minority Over-sampling Technique (SMOTE)	23
2.7.3 Mekanisme Kerja.....	24
2.7.4 Penerapan SMOTE pada Pemrosesan Teks (NLP).....	24
2.8 Text Preprocessing (Prapemrosesan Teks)	25
2.9 Metrik Evaluasi Kinerja Model	27
2.9.1 Confusion Matrix	28
2.9.2 Akurasi (Accuracy)	28
2.9.3 Presisi (Precision)	29
2.9.4 Recall (Sensitivitas)	29
2.9.5 F1-Score	29

BAB III TAHAPAN PELAKSANAAN.....	30
3.1 Tahapan Penelitian.....	30
3.1.1 Alur (a) Pelatihan Model	31
3.1.2 Alur (b) Pengujian Model	32
3.1.3 Alur (c) Implementasi Aplikasi.....	33
3.2 Pengumpulan Data	33
3.2.1 Alat dan Konfigurasi <i>Scraping</i>	34
3.2.2 Ekstraksi dan Transformasi Data	35
3.3 Text Preprocessing (Prapemrosesan Teks)	36
3.4 Pelabelan Data	41
3.5 Pemodelan dan Pelatihan IndoBERT.....	45
3.5.1 Ekstraksi Fitur dengan IndoBERT	45
3.5.2 Klasifikasi dengan MLPClassifier	48
3.5.3 Optimasi Hyperparameter dengan GridSearchCV.....	49
3.5.4 Perhitungan Mekanisme 3-Fold Cross-Validation.....	50
3.5.5 Skenario Eksperimen Pelatihan	51
3.6 Implementasi SMOTE untuk Ketidakseimbangan Data	52
3.6.1 Tahapan Implementasi SMOTE	53
3.6.2 Mekanisme Pembentukan Sampel Sintetis	53
3.6.3 Ilustrasi Proses Sintesis Data Menggunakan SMOTE	54
3.6.4 Implementasi SMOTE pada Embedding IndoBERT	56
3.7 Evaluasi Model	57
3.7.1 Confusion Matrix	57
3.7.2 Classification Report.....	59
3.7.3 Implementasi Streamlit	60
3.8 Perancangan Skenario Pengujian.....	61
BAB IV HASIL DAN PEMBAHASAN	62
4.1 Hasil Implementasi Streamlit.....	62
4.1.1 Tampilan Antarmuka Utama.....	63
4.1.2 Tampilan Demo Prediksi	64
4.2 Hasil Evaluasi Model	66
4.2.1 Hasil Pengujian Model Sebelum Optimasi Hyperparameter	66

4.2.2 Hasil Optimasi Hyperparameter Menggunakan GridSearchCV	68
4.2.3 Hasil Pengujian Model Setelah Penerapan SMOTE	70
4.2.4 Perbandingan Hasil Pengujian Model.....	71
4.3 Pembahasan.....	73
BAB V PENUTUP	76
5.1 Kesimpulan	76
5.2 Saran	77
DAFTAR PUSTAKA.....	78
DAFTAR RIWAYAT HIDUP	82



DAFTAR GAMBAR

Gambar 2.1 Bagan Cabang NLP [22]	10
Gambar 2.2 Bagan alur fase pemrosesan bahasa [22]	12
Gambar 2.3 Perbandingan Sentimen dan Deteksi Sikap.....	13
Gambar 2.4 Diagram Arsitektur Transformer [31]	17
Gambar 2.5 Ilustrasi Arsitektur BERT [31].....	18
Gambar 2.6 Tabel Confusion Matrix[12]	28
Gambar 3.1 Alur Penelitian	30
Gambar 3.2 Bagan Alir Preprocessing/Prapemrosesan Teks.....	37
Gambar 3.3 Bagan Alir Pelabelan	41
Gambar 3.4 Grafik Batang Distribusi Kelas Sebelum SMOTE.....	44
Gambar 3.5 Ekstraksi Fitur IndoBERT	45
Gambar 3.6 Parameter GridSearchCV.....	50
Gambar 3.7 Confusion Matrix	57
Gambar 3.8 Wireframe Implementasi dengan Streamlit	60
Gambar 3.9 Perancangan Skenario Pengujian.....	61
Gambar 4.1 Proses Pembuatan Streamlit.....	62
Gambar 4.2 Hasil Tampilan Streamlit	63
Gambar 4.3 Hasil Demo Prediksi Kasus 1.....	64
Gambar 4.4 Hasil Prediksi Kasus 2	65
Gambar 4.5 Confusion Matrix Sebelum Optimasi Hyperparameter.....	66
Gambar 4.6 Hasil Evaluasi Sebelum Optimasi.....	67
Gambar 4.7 Confusion Matrix Setelah Optimasi Hyperparameter	68
Gambar 4.8 Hasil Evaluasi Model Setelah Optimasi.....	69
Gambar 4.9 Confusion Matrix Model Setelah Penerapan SMOTE.....	70
Gambar 4.10 Hasil Evaluasi Model Setelah Penerapan SMOTE	71
Gambar 4.11 Hasil Perbandingan Pengujian Model.....	72

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu	7
Tabel 2.2 Perbandingan Sentiment Analysis dan Stance Detection [6].....	14
Tabel 2.3 Sampel Teks dari Dataset Uji Semeval 2016 [25]	15
Tabel 2.4 Varian dan Jenis Model IndoBERT [35]	21
Tabel 2.5 Contoh Hasil Data Cleaning [7].....	25
Tabel 3.1 Informasi Dataset.....	34
Tabel 3.2 Data Mentah (Raw Data) Komentar TikTok	35
Tabel 3.3 Hasil Cleaning	38
Tabel 3.4 Hasil Case Folding.....	38
Tabel 3.5 Entri Kamus Normalisasi Slang.....	39
Tabel 3.6 Hasil Tokenisasi.....	39
Tabel 3.7 Entri Kamus Stopword Removal	40
Tabel 3.8 Hasil Stemming	40
Tabel 3.9 Penyusunan Kamus Kata Kunci	41
Tabel 3.10 Penentuan Skema Klasifikasi 2 Kelas	42
Tabel 3.11 Contoh Hasil Pelabelan.....	43
Tabel 3.12 Distribusi Kelas Hasil Pelabelan.....	43
Tabel 3.13 Penjelasan Konfigurasi Tokenisasi	46
Tabel 3.14 Proses Tokenisasi IndoBERT pada Komentar K1.....	47
Tabel 3.15 Penjelasan Konfigurasi MLPClassifier	49
Tabel 3.16 Penjelasan Konfigurasi Grid Pencarian Hyperparameter	49
Tabel 3.17 Ilustrasi Skor F1-Macro per Fold (Kombinasi Hyperparameter Terbaik)	51
Tabel 3.18 Perbandingan Skenario Pelatihan	52
Tabel 3.19 Data Awal Vektor Embedding (4 dimensi Representatif)	54
Tabel 3.20 Perhitungan Sintesis Sampel Baru K6 Menggunakan Persamaan 2 ($\delta = 0.4$) ...	54
Tabel 3.21 Distribusi Kelas Sebelum dan Setelah Penerapan SMOTE	55
Tabel 3.22 Perhitungan Sintesis Sampel Baru pada 10 Dimensi Pertama Vektor IndoBERT ($\delta = 0.6$).....	56
Tabel 4.1 Konfigurasi Hyperparameter	68
Tabel 4.2 Perbandingan Hasil Pengujian Model.....	72