

**IMPLEMENTASI INDOBERT DAN INDOBERTTWEET
UNTUK DETEKSI UJARAN KEBENCIAN DI MEDIA
SOSIAL**



SKRIPSI

**EDWARD LEONARDO KOSASIH
00000055915**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**IMPLEMENTASI INDOBERT DAN INDOBERTTWEET
UNTUK DETEKSI UJARAN KEBENCIAN DI MEDIA
SOSIAL**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**EDWARD LEONARDO KOSASIH
0000055915**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Edward Leonardo Kosasih
Nomor Induk Mahasiswa : 00000055915
Program Studi : Informatika

Skripsi dengan judul:

Implementasi IndoBERT dan IndoBERTweet untuk Deteksi Ujaran Kebencian di Media Sosial

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 2 Juli 2025



(Edward Leonardo Kosasih)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI INDOBERT DAN INDOBERTWEET UNTUK DETEKSI UJARAN KEBENCIAN DI MEDIA SOSIAL

oleh

Nama : Edward Leonardo Kosasih
NIM : 00000055915
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

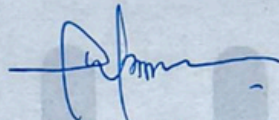
Telah diujikan pada hari Senin, 14 Juli 2025

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

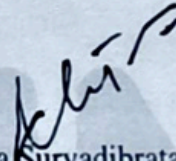
Ketua Sidang



(Sy. Yuliani Yakub, S.Kom., M.T., Ph.D)

NIDN: 0411037904

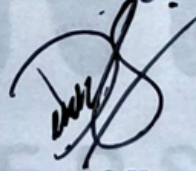
Penguji



(Alethea Suryadibrata, S.Kom., M.Eng.)

NIDN: 0322099201

Pembimbing



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Edward Leonardo Kosasih
NIM : 00000055915
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Implementasi IndoBERT dan
IndoBERTweet untuk Deteksi Ujaran
Kebencian di Media Sosial

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 2 Juli 2025

Yang menyatakan

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Edward Leonardo Kosasih

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji dan syukur dipanjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan petunjuk-Nya sehingga skripsi yang berjudul “Implementasi IndoBERT dan IndoBERTweet untuk Deteksi Ujaran Kebencian di Media Sosial” dapat diselesaikan dengan baik dan lancar.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana. Penyusunannya tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini disampaikan rasa terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Diharapkan hasil penelitian ini dapat memberikan kontribusi bagi pengembangan teknologi pemrosesan bahasa alami, khususnya dalam upaya deteksi ujaran kebencian di media sosial. Penelitian ini diharapkan tidak hanya bermanfaat bagi kalangan akademik, tetapi juga dapat dijadikan acuan dalam membangun sistem.

Tangerang, 2 Juli 2025

Edward Leonardo Kosasih

IMPLEMENTASI INDOBERT DAN INDOBERTTWEET UNTUK DETEKSI UJARAN KEBENCIAN DI MEDIA SOSIAL

Edward Leonardo Kosasih

ABSTRAK

Penggunaan media sosial yang tinggi di Indonesia meningkatkan risiko penyebaran ujaran kebencian secara masif, khususnya pada periode pemilihan umum. Ujaran kebencian pada platform seperti Twitter sering menasar kelompok rentan dan dapat memicu konflik sosial, memperkuat stereotip negatif, serta mengancam kohesi masyarakat. Oleh karena itu, diperlukan sistem deteksi otomatis yang akurat dan kontekstual. Pendekatan berbasis Transformer mulai dimanfaatkan karena kemampuannya dalam menangkap makna secara kontekstual. Penelitian ini membandingkan dua model pra-latih berbasis Transformer untuk Bahasa Indonesia, yaitu IndoBERT dan IndoBERTweet, dalam mendeteksi ujaran kebencian. Kedua model menjalani proses *fine-tuning* menggunakan data pelatihan dan diuji menggunakan testing dataset. Evaluasi dilakukan dengan menggunakan metrik F1-score dan akurasi. Hasil pengujian menunjukkan bahwa IndoBERTweet memiliki performa lebih baik (F1-score 0,8651; akurasi 88,38%) dibandingkan IndoBERT (F1-score 0,85; akurasi 87,47%). Temuan ini menunjukkan pentingnya pemilihan model yang sesuai dengan domain data untuk membangun sistem deteksi ujaran kebencian yang lebih akurat.

Kata kunci: Deteksi Ujaran Kebencian, IndoBERT, IndoBERTweet, Media Sosial, Twitter



IMPLEMENTATION OF INDOBERT AND INDOBERTTWEET FOR HATE SPEECH DETECTION ON SOCIAL MEDIA

Edward Leonardo Kosasih

ABSTRACT

Keywords: *Hate Speech Detection, IndoBERT, IndoBERTweet, Social Media, Twitter* The high usage of social media in Indonesia increases the risk of widespread dissemination of hate speech, particularly during election periods. Hate speech on platforms such as Twitter often targets vulnerable groups and can trigger social conflict, reinforce negative stereotypes, and threaten societal cohesion. Therefore, an automatic detection system that is both accurate and context-aware is urgently needed. Transformer-based approaches have begun to be utilized due to their ability to capture contextual meaning. This study compares two pre-trained Transformer-based models for the Indonesian language—IndoBERT and IndoBERTweet—in detecting hate speech. Both models underwent a fine-tuning process using training data and were evaluated using a testing dataset. The evaluation employed F1-score and accuracy as metrics. The results show that IndoBERTweet outperforms IndoBERT (F1-score: 0.8651; accuracy: 88.38)

Keywords: *Hate Speech Detection, IndoBERT, IndoBERTweet, Social Media, Twitter*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.1 Ujaran Kebencian	6
2.1.1 Definisi Ujaran Kebencian	6
2.1.2 Jenis-Jenis Ujaran Kebencian	6
2.1.3 Dampak Ujaran Kebencian	7
2.2 Media Sosial	7
2.3 Natural Language Processing (NLP)	8
2.4 Bidirectional Encoder Representations from Transformers (BERT)	8
2.4.1 Arsitektur BERT	8
2.4.2 Pre-training Tasks	8
2.4.3 Tokenisasi dan Representasi Input	9
2.4.4 Fine-tuning BERT	10
2.4.5 Keunggulan BERT	10
2.5 IndoBERT	11
2.6 IndoBERTweet	11
2.7 Pra-pemrosesan Teks	11
2.8 Evaluation Metrics	12
BAB 3 METODOLOGI PENELITIAN	14
3.1 Gambaran Umum Penelitian	14
3.1.1 Spesifikasi Perangkat	15
3.2 Studi Literatur	15
3.3 Pengumpulan Data	15
3.4 Pengolahan Data	16
3.5 Perancangan Model	18
3.5.1 IndoBERT Dan IndoBERTweet	20
3.6 Testing	21
3.7 Dokumentasi	21
BAB 4 HASIL DAN DISKUSI	22

4.1	Hasil Implementasi Pembelajaran Mesin	22
4.1.1	Pengumpulan Dataset	22
4.1.2	Pengolahan Data	23
4.1.3	Perancangan IndoBERT dan IndoBERTweet	25
4.2	Hasil Uji Coba	31
4.2.1	Pengujian Hyperparameter	31
4.2.2	Pengujian Model Terbaik	33
4.2.3	Diskusi	36
BAB 5	SIMPULAN DAN SARAN	48
5.1	Simpulan	48
5.2	Saran	48
DAFTAR PUSTAKA		49



DAFTAR TABEL

Tabel 4.1	Kombinasi Batch Size, Learning Rate, dan Epoch	31
Tabel 4.2	Hasil Pengujian <i>Hyperparameter</i> IndoBERT Terbaik	32
Tabel 4.3	Hasil Pengujian <i>Hyperparameter</i> IndoBERTweet Terbaik .	33
Tabel 4.4	Kombinasi Parameter Terbaik Untuk Model IndoBERT dan IndoBERTweet	34
Tabel 4.5	Hasil Evaluasi Model IndoBERT dengan Variasi Batch . .	37
Tabel 4.6	F1-Score Model IndoBERT pada Variasi Nilai Learning Rate	38
Tabel 4.7	F1-Score Model IndoBERT per Epoch	40
Tabel 4.8	Hasil Evaluasi Model IndoBERTweet dengan Variasi Batch	41
Tabel 4.9	F1-Score Model IndoBERTweet pada Variasi Nilai Learning Rate	43
Tabel 4.10	<i>accuracy</i> dan F1-Score Model IndoBERTweet per Epoch pada	44
Tabel 5.1	Hasil Evaluasi Model IndoBERT berdasarkan Kombinasi Batch Size, Learning Rate, dan Epoch	61
Tabel 5.2	Hasil Evaluasi Model IndoBERTweet berdasarkan Kombinasi Batch Size, Learning Rate, dan Epoch	63

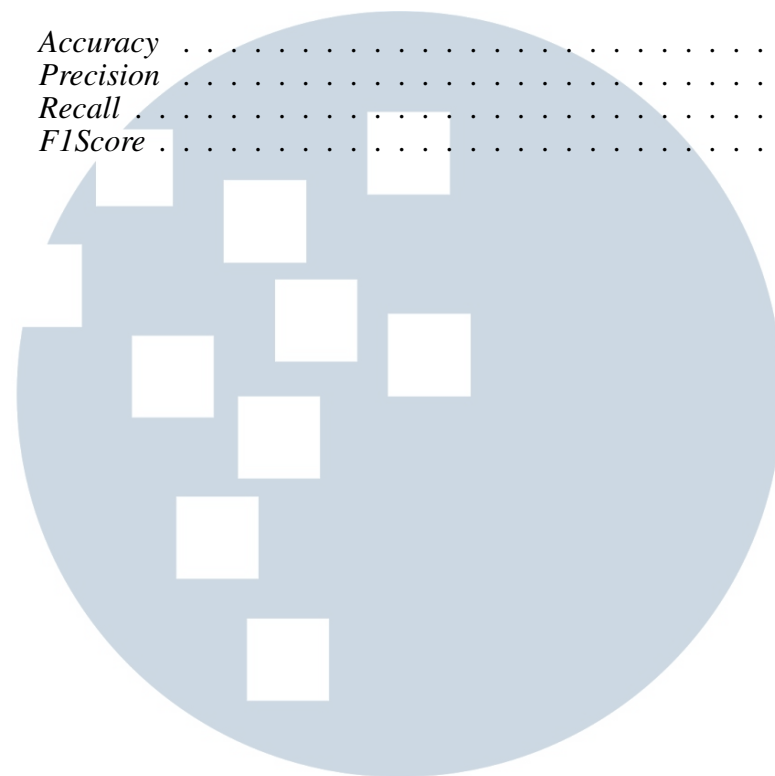


DAFTAR GAMBAR

Gambar 2.1	Gambaran Pre-training BERT dan Fine-Tuning BERT [1]	9
Gambar 2.2	Representasi input pada BERT, terdiri dari token embeddings, segment embeddings, dan position embeddings [1].	10
Gambar 3.1	<i>Flowchart</i> alur Keseluruhan Penelitian	14
Gambar 3.2	Potongan Data <i>Hate speech</i>	16
Gambar 3.3	<i>Flowchart</i> Pengelolaan Data	17
Gambar 3.4	<i>Flowchart</i> Menghapus Karakter Tidak Perlu	18
Gambar 3.5	<i>Flowchart</i> Memperbaiki Kata Salah Ejaan	18
Gambar 3.6	<i>Flowchart</i> Implementasi Model	19
Gambar 3.7	<i>Flowchart</i> Training IndoBERT dan IndoBERTweet	20
Gambar 3.8	<i>Flowchart</i> Testing	21
Gambar 4.1	Kode Memasukan Data <i>Hate Speech</i>	22
Gambar 4.2	Hasil Pembacaan Data <i>Hate Speech</i>	23
Gambar 4.3	Kode <i>pre-processing</i> data <i>Hate Speech</i>	24
Gambar 4.4	Testing Kode <i>Pre-processing</i> Data	24
Gambar 4.5	Contoh Data Sebelum <i>Pre-Processing</i>	24
Gambar 4.6	Contoh Data Sesudah <i>Pre-Processing</i>	25
Gambar 4.7	Tokenisasi Dataset Tweet	25
Gambar 4.8	Potongan Kode Fungsi TweetDataset	26
Gambar 4.9	Potongan Kode Penggabukan tokens dan labels	26
Gambar 4.10	Potongan Kode Fungsi Pembagian Dataset IndoBERT	27
Gambar 4.11	Potongan Kode Fungsi Pembagian Dataset IndoBERTweet	27
Gambar 4.12	Potongan Kode Impor Model IndoBERT dan IndoBERTweet	28
Gambar 4.13	Potongan Kode Evaluasi	29
Gambar 4.14	Potongan Kode Memasukan <i>Hyperparameter</i> Pada Model Pelatihan	30
Gambar 4.15	Potongan Kode Pelatihan	31
Gambar 4.16	Nilai True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) pada model IndoBERT	34
Gambar 4.17	Nilai True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) pada model IndoBERTweet	35
Gambar 4.18	Re-presentasi Tabel 4.5 dalam Bentuk Grafik Plot	37
Gambar 4.19	Re-presentasi Tabel 4.6 dalam Bentuk Grafik Plot	39
Gambar 4.20	Re-presentasi Tabel 4.7 dalam Bentuk Grafik Plot	40
Gambar 4.21	Re-presentasi Tabel 4.8 dalam Bentuk Grafik Plot	42
Gambar 4.22	Re-presentasi Tabel 4.9 dalam Bentuk Grafik Plot	43
Gambar 4.23	Re-presentasi Tabel 4.10 dalam Bentuk Grafik Plot	45
Gambar 4.24	Nilai True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) untuk kelas <i>negative</i>	45
Gambar 4.25	Nilai True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) pada model IndoBERTweet untuk kelas <i>negative</i>	47

DAFTAR RUMUS

Rumus 2.1	<i>Accuracy</i>	12
Rumus 2.2	<i>Precision</i>	12
Rumus 2.3	<i>Recall</i>	12
Rumus 2.4	<i>F1Score</i>	12



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	52
Lampiran 2	Formulir Bimbingan	59
Lampiran 3	Hasil Uji Coba Setiap Hyperparameter IndoBERT	61
Lampiran 4	Hasil Uji Coba Setiap Hyperparameter IndoBERTweet	63

