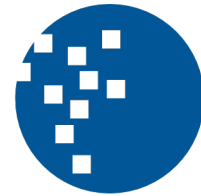


**PEMANFAATAN MODEL BLOOM UNTUK DETEKSI
UJARAN KEBENCIAN PADA TEKS CAMPURAN BAHASA
INDONESIA-INGGRIS**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

KEVIN SORENSEN
00000062002

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

**PEMANFAATAN MODEL BLOOM UNTUK DETEKSI
UJARAN KEBENCIAN PADA TEKS CAMPURAN BAHASA
INDONESIA-INGGRIS**



SKRIPSI

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

KEVIN SORENSEN

00000062002

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Kevin Sorensen

Nomor Induk Mahasiswa : 00000062002

Program Studi : Informatika

Skripsi dengan judul:

Pemanfaatan Model BLOOM untuk Deteksi Ujaran Kebencian pada Teks Campuran Bahasa Indonesia-Inggris

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, 26 Juni 2025



(Kevin Sorensen)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PEMANFAATAN MODEL BLOOM UNTUK DETEKSI UJARAN KEBENCIAN PADA TEKS CAMPURAN BAHASA INDONESIA-INGGRIS

oleh

Nama : Kevin Sorensen
NIM : 00000062002
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Rabu, 16 Juli 2025

Pukul 15.00 s/d 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji

(Alethea Suryadibrata, S.Kom., M.Eng.) (Eka Jaya Harsono, S.Kom., M.Eng.Sc.)

NIDN: 0322099201

NIDN: 8343771672130333

Pembimbing

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

NIDN: 0315109103

Ketua Program Studi Informatika,

(Assoc. Prof. 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 : Kevin Sorensen
NIM : 00000062002
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Pemanfaatan Model BLOOM untuk
Deteksi Ujaran Kebencian pada Teks
Campuran Bahasa Indonesia-Inggris

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, 26 Juni 2025

Yang menyatakan



Kevin Sorensen

**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

“We are not horses. We are humans. Humans are...”

— Seong Gi-hun (Player 456), *Squid Game*



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik dan lancar. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Dalam perjalanan menyelesaikan tugas ini, penulis menerima banyak dukungan, bantuan, dan semangat dari berbagai pihak, yang sangat berarti dan tak ternilai.

Dengan penuh rasa hormat dan terima kasih, penulis menyampaikan apresiasi 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 UMN.
3. Bapak Assoc. Prof. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika sekaligus pembimbing Tugas Akhir, atas bimbingan, masukan, dan arahannya.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., selaku Sekretaris Program Studi Informatika, yang telah mendukung dan membantu kelancaran penulis selama menjalani masa studi.
5. Seluruh dosen dan staf Informatika UMN yang telah membekali penulis dengan ilmu dan pengalaman berharga selama perkuliahan.
6. Kedua orang tua dan keluarga tercinta atas doa, kasih sayang, serta dukungan moral dan material yang tiada henti.
7. Teman-teman di Himpunan Mahasiswa Informatika (HMIF) UMN, khususnya divisi Research and Development, atas kontribusi dan kebersamaannya.
8. Rekan-rekan seperjuangan, sahabat “Cina berkeliaran”, “Anak Baik”, “Manusia Kuat”, “Paguyuban Maba Biasa”, serta semua pihak yang telah memberi warna, tawa, dan semangat dalam perjalanan ini.

9. Teman-teman pengguna ChatSorensen yang telah menjadi teman diskusi, berbagi tawa, dan semangat selama masa perkuliahan.
10. Seseorang yang sangat berarti dalam hidup penulis atas kesabaran, dukungan, dan motivasi yang tulus sepanjang proses ini.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki kekurangan, oleh karena itu masukan dan saran yang membangun sangat diharapkan. Semoga karya ini dapat memberikan manfaat dan menjadi referensi untuk pengembangan lebih lanjut.

Tangerang, 26 Juni 2025



Kevin Sorensen

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

PEMANFAATAN MODEL BLOOM UNTUK DETEKSI UJARAN KEBENCIAN PADA TEKS CAMPURAN BAHASA INDONESIA-INGGRIS

Kevin Sorensen

ABSTRAK

Ujaran kebencian merupakan masalah krusial dalam komunikasi digital, terutama pada konteks campuran bahasa (code-mixing) antara Bahasa Indonesia dan Bahasa Inggris. Sistem deteksi konvensional sering gagal mengidentifikasi ujaran kebencian dalam teks campuran karena keterbatasan data dan model yang bersifat monolingual. Penelitian ini mengkaji kinerja model BLOOM, yaitu model bahasa berbasis transformer multibahasa, dalam mendeteksi ujaran kebencian pada teks campuran. Model dilatih menggunakan dataset yang diperluas hingga 30.000 sampel yang mencakup Bahasa Indonesia, Bahasa Inggris, dan teks campuran. Hasil pelatihan menunjukkan akurasi validasi sebesar 99,97% dan nilai F1-score sebesar 99,97%. Evaluasi pada dataset yang beragam menunjukkan ketangguhan model serta mengidentifikasi tantangan, terutama dalam mendeteksi ekspresi informal dan kontekstual. Hasil ini menunjukkan bahwa model multibahasa seperti BLOOM memiliki potensi besar dalam menangani deteksi ujaran kebencian pada lingkungan teks campuran.

Kata kunci: code-mixing, deteksi ujaran kebencian, model BLOOM, model multibahasa, teks Indonesia-Inggris



LEVERAGING THE BLOOM MODEL FOR ENHANCED DETECTION OF HATE SPEECH IN INDONESIAN-ENGLISH CODE-MIXED TEXTS

Kevin Sorensen

ABSTRACT

Hate speech detection is a critical task for fostering inclusive communication in digital platforms, especially in code-mixing scenarios where linguistic boundaries blur. Traditional detection systems often falter when processing Indonesian-English mixed-language text due to their reliance on monolingual datasets and models. This study evaluates the performance of the BLOOM model, a multilingual transformer-based language model, in identifying hate speech within code-mixed text. Using an expanded dataset of 30,000 samples encompassing Indonesian, English, and mixed-language entries, the model was fine-tuned, achieving a validation accuracy of 99.97% and an F1-score of 99.97%. Evaluation on diverse datasets demonstrated the model's robustness and highlighted areas requiring improvement, particularly in detecting informal and context-dependent expressions. These findings affirm the potential of multilingual models like BLOOM in addressing the intricacies of hate speech detection in code-mixed environments, offering new avenues for future advancements.

Keywords: Bloom model, code-mixing, hate speech detection, Indonesian-English text, multilingual models



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	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	4
2.1 Studi Literatur	4
2.2 Model BLOOM	9
BAB 3 METODOLOGI PENELITIAN	11
3.1 Pengumpulan Data	11
3.2 Pra-pemrosesan Data	12
3.3 Fine-Tuning	13
3.3.1 Pembagian Data	14
3.3.2 Pelatihan dan Validasi	14
3.4 Penyetelan Hiperparameter	21
3.5 Sumber Daya Komputasi	30
3.5.1 Fine-Tuning	30
3.5.2 Penyetelan Hiperparameter	31
BAB 4 HASIL DAN DISKUSI	33
4.1 Hasil Penelitian	33
4.1.1 Performa Model pada Dataset Uji Utama (Data Primer)	33
4.1.2 Performa Model pada Dataset Uji Khusus (Data Independensi)	34
4.1.3 Performa Model pada Dataset Modifikasi (Bahasa Tidak Formal)	36
4.2 Analisis Komparatif Performa Model	37
BAB 5 SIMPULAN DAN SARAN	39
5.1 Simpulan	39
5.2 Saran	39
DAFTAR PUSTAKA	40

DAFTAR TABEL

Tabel 2.1	Analisis komparatif model NLP untuk deteksi ujaran kebencian	6
Tabel 3.1	Hasil metrik Fine-Tuning selama 11 epoch	15
Tabel 3.2	Hasil Penyesuaian Hyperparameter Selama Beberapa Epoch untuk Berbagai Ukuran Batch dan Learning Rate . .	22
Tabel 3.3	Metrik kinerja selama beberapa epoch	24
Tabel 4.1	Contoh teks salah klasifikasi dari dataset uji khusus	36



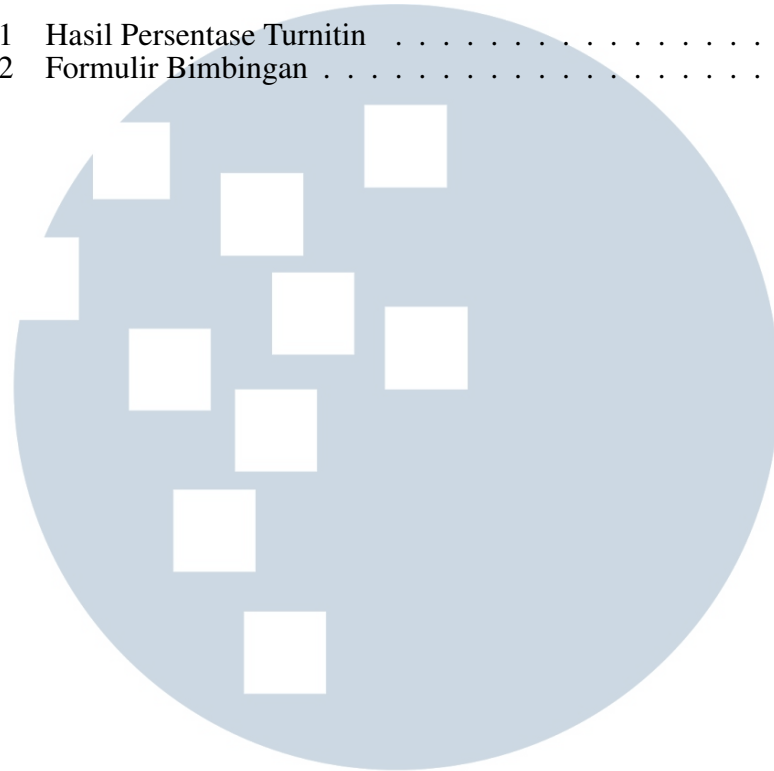
DAFTAR GAMBAR

Gambar 2.1	Arsitektur model BLOOM	9
Gambar 3.1	Pra-pemrosesan data	12
Gambar 3.2	Loss pelatihan dan validasi selama 11 epoch	17
Gambar 3.3	Akurasi pelatihan dan validasi selama 11 epoch	18
Gambar 3.4	Presisi pelatihan dan validasi selama 11 epoch	19
Gambar 3.5	Recall pelatihan dan validasi selama 11 epoch	20
Gambar 3.6	F1-Score pelatihan dan validasi selama 11 epoch	21
Gambar 3.7	Loss pelatihan dan validasi selama 20 epoch	26
Gambar 3.8	Akurasi pelatihan dan validasi selama 20 epoch	27
Gambar 3.9	Presisi pelatihan dan validasi selama 20 epoch	28
Gambar 3.10	Recall pelatihan dan validasi selama 20 epoch	29
Gambar 3.11	F1-Score pelatihan dan validasi selama 20 epoch	30
Gambar 4.1	Heatmap Confusion Matrix pada dataset uji utama	34
Gambar 4.2	Heatmap Confusion Matrix pada dataset uji khusus	35
Gambar 4.3	Heatmap Confusion Matrix pada dataset uji penulis yang dimodifikasi	37
Gambar 4.4	Perbandingan performa model pada tiga skenario pengujian	38



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	42
Lampiran 2	Formulir Bimbingan	43



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA