

**IMPLEMENTASI MODEL BI-LSTM MULTIBAHASA
UNTUK DETEKSI BERITA HOAKS DALAM BAHASA
INDONESIA DAN MELAYU**



LAPORAN MBKM PENELITIAN

**ANDREW THOMAS AGUSTINUS
00000059999**

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

**IMPLEMENTASI MODEL BI-LSTM MULTIBAHASA
UNTUK DETEKSI BERITA HOAKS DALAM BAHASA
INDONESIA DAN MELAYU**



**ANDREW THOMAS AGUSTINUS
00000059999**

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

HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Andrew Thomas Agustinus

NIM : 00000059999

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan MBKM Penelitian saya yang berjudul:

Implementasi Model Bi-LSTM Multibahasa untuk Deteksi Berita Hoaks dalam Bahasa Indonesia dan Melayu

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 23 Juni 2025



(Andrew Thomas Agustinus)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Andrew Thomas Agustinus
NIM : 00000059999
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Laporan MBKM Penelitian

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- ☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 23 Juni 2025

Yang menyatakan



Andrew Thomas Agustinus

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

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

Halaman Persembahan / Motto

"No temptation has overtaken you except what is common to mankind. And God is faithful; he will not let you be tempted beyond what you can bear. But when you are tempted, he will also provide a way out so that you can endure it."

1 Corinthians 10:13 NIV

"God would never put us through all this suffering if He didn't think we could bear it"

Konno Yuuki (Zekken)



KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan MBKM Penelitian ini dengan judul: "Implementasi model Bi-LSTM multibahasa untuk deteksi berita hoaks dalam bahasa Indonesia dan Melayu" dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

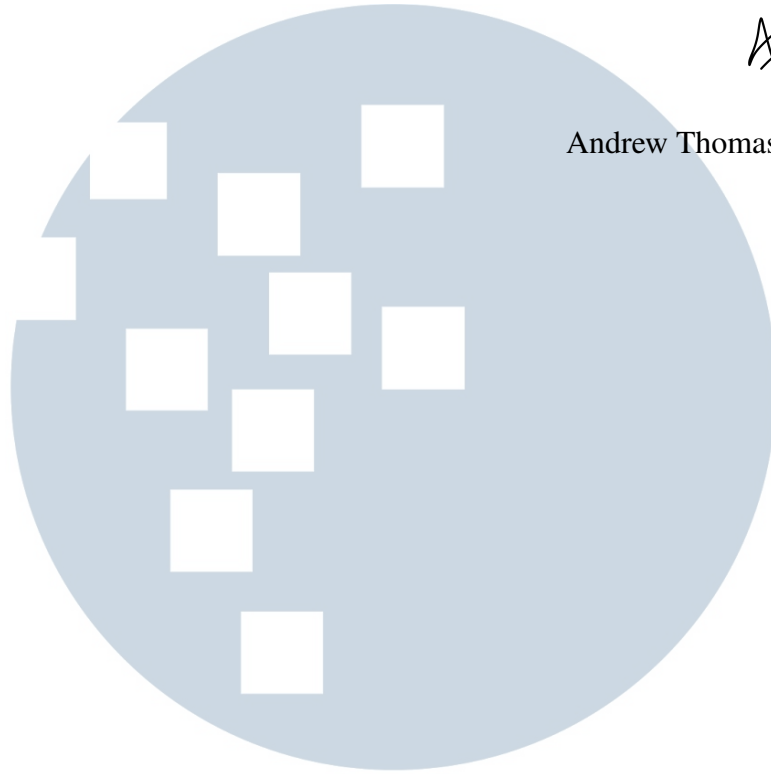
1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Alethea Suryadibrata, S.Kom., M.Eng., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya laporan penelitian ini.
5. Ibu Dr. Sy. Yuliani Yakub, S.Kom., M.T, selaku Kepala Tim Peneliti UNIIC UMN yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas penelitian yang dilakukan.
6. Sabrina sebagai teman di Tim Peneliti UNIIC UMN yang membantu dengan penelitian ini.
7. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan penelitian ini.

Semoga karya ilmiah ini dapat memberikan manfaat yang baik sebagai sumber informasi maupun inspirasi seluruh kalangan.

Tangerang, 23 Juni 2025



Andrew Thomas Agustinus



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

IMPLEMENTASI MODEL BI-LSTM MULTIBAHASA UNTUK DETEKSI BERITA HOAKS DALAM BAHASA INDONESIA DAN MELAYU

Andrew Thomas Agustinus

ABSTRAK

Penyebaran hoaks di Indonesia dan Malaysia menjadi ancaman serius terhadap kepercayaan publik dan keamanan siber nasional. Rendahnya literasi digital, tingginya penetrasi media sosial, serta keragaman bahasa menjadikan kawasan ini sangat rentan terhadap disinformasi. Untuk menjawab tantangan tersebut, penelitian ini mengusulkan model deteksi hoaks multibahasa berbasis *Bidirectional Long Short-Term Memory* (Bi-LSTM) untuk bahasa Indonesia dan Melayu. Dataset terdiri dari 30.343 data Indonesia dan 37.592 data Melayu yang dipra-pemrosesan melalui pembersihan teks, tokenisasi, penghapusan *stopwords*, *stemming*, dan *padding*. Model dilatih menggunakan arsitektur embedding, dua lapisan Bi-LSTM, serta *dropout*, dengan *Adam optimizer* dan fungsi kehilangan *binary crossentropy*. Hasil pelatihan menunjukkan peningkatan akurasi dari 81% menjadi 99,5%, serta penurunan *loss* dari 0,3948 menjadi 0,0161 dalam 10 *epoch*. Evaluasi lebih lanjut menunjukkan akurasi 97,17% pada data pelatihan, 91,68% pada validasi, dan 92,30% pada pengujian. Nilai presisi, *recall*, dan F1-score seimbang pada kedua kelas, tanpa indikasi bias. Hasil ini menunjukkan bahwa Bi-LSTM efektif dalam mengenali karakteristik semantik dan sintaktik berita hoaks maupun asli dalam konteks multibahasa secara akurat dan andal.

Kata kunci: Bi-LSTM, Deteksi Hoaks, Keamanan Siber, Multibahasa, Pemrosesan Bahasa Alami

U M N
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

**MULTILINGUAL BI-LSTM MODEL IMPLEMENTATION FOR HOAX
NEWS DETECTION IN INDONESIAN AND MALAY**

Andrew Thomas Agustinus

ABSTRACT

The spread of hoaxes in Indonesia and Malaysia poses a serious threat to public trust and national cybersecurity. Low digital literacy, high social media penetration, and linguistic diversity make the region vulnerable to online disinformation. To address this issue, this study proposes a multilingual hoax detection model based on a Bidirectional Long Short-Term Memory (Bi-LSTM) architecture for the Indonesian and Malay languages. The dataset consists of 30,343 Indonesian and 37,592 Malay news texts, preprocessed through text cleaning, tokenization, stopword removal, stemming, and padding. The model architecture includes embedding layers, two Bi-LSTM layers, and dropout, trained using the Adam optimizer and binary crossentropy loss function. Training results show an accuracy increase from 81% to 99.5% and a decrease in training loss from 0.3948 to 0.0161 over 10 epochs. Further evaluation indicates 97.17% accuracy on training data, 91.68% on validation, and 92.30% on test data. The model achieves balanced precision, recall, and F1-score for both classes, with no significant bias. These findings confirm that the Bi-LSTM model effectively learns semantic and syntactic patterns for accurate multilingual hoax detection.

Keywords: *Bi-LSTM, Cybersecurity, Hoax Detection, Multilingual, Natural Language Processing*



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
HALAMAN PERSEMBAHAN/MOTO	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Urgensi Penelitian	3
1.6 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Pengertian Hoaks	5
2.1.1 Sumber Dataset	5
2.2 Text Preprocessing	6
2.2.1 Pembersihan Teks	6
2.2.2 Tokenisasi	7
2.2.3 Tokenisasi dan Padding	7
2.3 Long Short-Term Memory (LSTM)	8
2.3.1 Arsitektur LSTM	8
2.3.2 Formulasi Matematis	9
2.4 Bidirectional LSTM (Bi-LSTM)	9
2.4.1 Arsitektur Bi-LSTM	9
2.4.2 Embedding	10
2.5 Evaluasi Metrik	10
2.5.1 Confusion Matrix	10
2.5.2 Metrik Klasifikasi Utama	11
2.5.3 Classification Report	11
BAB 3 METODOLOGI PENELITIAN	12
3.1 Pengumpulan Data	12
3.1.1 Dataset Malaysia	13
3.1.2 Dataset Indonesia	13
3.2 Data Cleaning	13
3.3 Pra-pemrosesan Teks	15
3.3.1 Tokenisasi	15
3.3.2 Penghapusan Stopwords	16
3.3.3 Stemming	16
3.4 Split Data	17
3.5 Tokenisasi dan Padding	18
3.6 Konfigurasi Model	19

3.6.1	Arsitektur Model BiLSTM	19
3.6.2	Konfigurasi Pelatihan Model	20
3.7	Evaluasi Model	20
3.8	Pembuatan Laporan dan Dokumentasi	21
BAB 4	HASIL DAN PEMBAHASAN	22
4.1	Hasil	22
4.1.1	Performa Pelatihan dan Validasi	22
4.1.2	Evaluasi Klasifikasi pada Data Pelatihan, Validasi, dan Uji	22
4.1.3	Confusion Matrix	24
4.1.4	Hasil Prediksi Berita	26
4.2	Pembahasan	29
4.2.1	Analisis Performa dan Evaluasi Model	29
BAB 5	SIMPULAN DAN SARAN	31
5.1	Simpulan	31
5.2	Saran	31
DAFTAR PUSTAKA		33



DAFTAR TABEL

Tabel 2.1	Struktur Matriks Konfusi	10
Tabel 3.1	Distribusi Data Dataset Hoaks Berbahasa Melayu dan Indonesia	13
Tabel 3.2	Konfigurasi Hyperparameter Pelatihan	20
Tabel 4.1	Hasil Pelatihan dan Validasi Tiap Epoch	22
Tabel 4.2	Classification Report pada Data Pelatihan	23
Tabel 4.3	Classification Report pada Data Validasi	23
Tabel 4.4	Classification Report pada Data Uji	23



DAFTAR GAMBAR

Gambar 3.1	Arsitektur metode penelitian	12
Gambar 4.1	Confusion matrix pada data pelatihan	24
Gambar 4.2	Confusion matrix pada data validasi	25
Gambar 4.3	Confusion matrix pada data uji	25
Gambar 4.4	Prediksi berita hoaks dengan confidence 98.98%	28
Gambar 4.5	Prediksi berita asli dengan confidence 99.80%	28



DAFTAR KODE

Kode 3.1	Proses pembersihan data teks pada dataset Indonesia	14
Kode 3.2	Proses tokenisasi menggunakan NLTK	15
Kode 3.3	Penghapusan <i>stopwords</i> menggunakan Sastrawi	16
Kode 3.4	Proses stemming menggunakan StemmerFactory dari Sastrawi .	16
Kode 3.5	Pembagian data menjadi train, validation, dan test set	17
Kode 3.6	Proses tokenisasi dan padding pada data teks	18
Kode 3.7	Arsitektur Model BiLSTM untuk Klasifikasi Teks	19
Kode 4.1	Implementasi Antarmuka Prediksi Berita Hoaks	26



DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Research	36
Lampiran 2	MBKM-02 MBKM Research Card	37
Lampiran 3	MBKM-03 Daily Task - Research	38
Lampiran 4	MBKM-04 Verification Form of Report MBKM Research . . .	45
Lampiran 5	Form Bimbingan	46
Lampiran 6	Pengecekan Hasil Turnitin	47
Lampiran 7	Surat Penerimaan Penelitian	55
Lampiran 8	Surat Kerjasama Penelitian	56
Lampiran 9	Draft Paper	57

