



**RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX
MENGUNAKAN MODEL NLP IndoBERT**

TUGAS AKHIR

APLIKATIF

SAMUEL LAMTAMA PURBA

SAMUEL LAMTAMA PURBA
UNIVERSITAS
MERCU BUANA
41522010242

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026



**RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX
MENGUNAKAN MODEL NLP IndoBERT**

**TUGAS AKHIR
APLIKATIF**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana



SAMUEL LAMTAMA PURBA

41522010242

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2026

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : SAMUEL LAMTAMA PURBA

NIM : 41522010242

Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir berjudul:

“RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX MENGGUNAKAN MODEL NLP IndoBERT” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta, atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau hak pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Jakarta, 28 Agustus 2026



Samuel Lamtama Purba

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : SAMUEL LAMTAMA PURBA
NIM /Student ID Number : 41522010242
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

**“RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX
MENGUNAKAN MODEL NLP IndoBERT”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

24 Agustus 2026

dengan nilai persentase sebesar :

/with a percentage value of:

6%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer**
Universitas Mercu Buana. */declared to meet standards in accordance with applicable
regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1ibQOsGsp5uWsPFPMeK02UOWUu48og6rV/view?usp=drive_link



Jakarta, 24 Agustus 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto

Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : SAMUEL LAMTAMA PURBA
NIM : 41522010242
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Tugas Akhir : RANCANG BANGUN WEBSITE PENGECEKAN
BERITA HOAX MENGGUNAKAN MODEL NLP
IndoBERT

Telah berhasil dipertahankan pada sidang tanggal 28 Agustus 2026, dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan Oleh

Pembimbing



Saruni Dwiasnati, S.T., M.M., M.Kom

NIDN/NUPTK: 0325128802

Jakarta, 28 Agustus 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI

NIDN/NUPTK: 0320037002

Ketua Program Studi

Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom

NIDN/NUPTK: 0225067701

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Rancang Bangun Website Pengecekan Berita Hoax Menggunakan Model NLP IndoBERT”** sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Penelitian ini berangkat dari kepedulian penulis terhadap maraknya penyebaran berita hoax di media sosial dan platform digital Indonesia. Penulis percaya bahwa teknologi pemrosesan bahasa alami (*Natural Language Processing*) dapat membantu masyarakat mengidentifikasi berita palsu secara lebih cepat dan akurat, sekaligus menumbuhkan literasi digital. Untuk itu, penelitian ini tidak hanya melatih dan mengevaluasi model deteksi, tetapi juga membangun aplikasi web yang dapat diakses publik dengan autentikasi dua peran (pengguna dan administrator).

Berkat dukungan, motivasi, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Bapak Dr. Ir. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. **Ibu Saruni Dwiasnati, S.T., M.M., M.Kom.** selaku dosen pembimbing yang telah memberikan pengarahan, motivasi, serta meluangkan waktu, tenaga, dan pikiran selama penyusunan skripsi ini.
5. Kedua orang tua dan keluarga penulis yang selalu memberikan dukungan moral, materiil, dan doa selama menjalani masa studi.
6. Seluruh teman seperjuangan mahasiswa Teknik Informatika yang telah berbagi diskusi, masukan, dan semangat.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Akhir kata, penulis berharap semoga skripsi ini bermanfaat bagi pembaca serta memberikan kontribusi positif bagi pengembangan teknologi deteksi hoax dan literasi digital di Indonesia.

Jakarta, 28 Agustus 2026



Samuel Lamtama Purba



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Samuel Lamtama Purba
NIM : 41522010242
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika
Judul Laporan Skripsi : Rancang Bangun Website Pengecekan Berita Hoax Menggunakan Model NLP IndoBERT

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Agustus 2026

Yang menyatakan,



Samuel Lamtama Purba

RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX MENGUNAKAN MODEL NLP IndoBERT

SAMUEL LAMTAMA PURBA

ABSTRAK

Penyebaran berita hoax di media sosial mengancam kohesi sosial dan literasi digital masyarakat Indonesia, sementara pemeriksaan fakta secara manual terbatas dalam skala dan kecepatan. Penelitian ini merancang bangun aplikasi web pengecekan berita hoax berbahasa Indonesia (Checkhoax) yang memanfaatkan model NLP IndoBERT. Model IndoBERT (indobenchmark/indobert-base-p1) di-*fine-tune* pada dataset gabungan 28.160 sampel dari lima sumber dan mencapai akurasi 97,59% serta F1-score makro 0,9758 pada data uji *held-out*. Sistem dilengkapi tiga lapisan pendukung kepercayaan, yaitu kalibrasi *confidence* dengan *temperature scaling* ($T^*=2,1834$) yang menurunkan *Expected Calibration Error* dari 2,29% menjadi 1,04%; pencocokan semantik berbasis SBERT terhadap indeks 14.779 hoaks terverifikasi untuk menangkap hoaks yang ditulis ulang; serta penilaian kredibilitas sumber yang direkonsiliasi dengan prediksi model. Aplikasi dibangun dengan arsitektur tiga lapis (React, Express, dan Flask) serta autentikasi dua peran (pengguna dan administrator) berbasis JSON Web Token dan bcrypt, mendukung analisis teks, URL, dan gambar (OCR), *explainable AI*, integrasi pemeriksa fakta eksternal, dasbor statistik, dan riwayat pribadi. Sistem lulus 20 skenario pengujian *black-box* serta 41 uji otomatis tanpa galat. Penelitian ini menjawab lima celah, yaitu implementasi *end-to-end*, integrasi multi-dataset, kalibrasi *confidence*, pencocokan semantik, dan integrasi kredibilitas sumber, sekaligus memposisikan sistem sebagai alat bantu literasi digital yang akurat, transparan, dan dapat dipercaya.

Kata kunci: deteksi hoax, IndoBERT, confidence calibration, semantic matching, kredibilitas sumber, autentikasi, sistem web, NLP

RANCANG BANGUN WEBSITE PENGECEKAN BERITA HOAX MENGUNAKAN MODEL NLP IndoBERT

SAMUEL LAMTAMA PURBA

ABSTRACT

The spread of hoax news on social media threatens social cohesion and digital literacy in Indonesia, while manual fact-checking is limited in scale and speed. This research designs and builds an Indonesian hoax news checking web application (Checkhoax) that leverages the IndoBERT NLP model. IndoBERT (indobenchmark/indobert-base-p1) is fine-tuned on a combined dataset of 28,160 samples from five sources, achieving 97.59% accuracy and a 0.9758 macro F1-score on held-out test data. The system adds three trust layers: confidence calibration via temperature scaling ($T^=2.1834$), which lowers the Expected Calibration Error from 2.29% to 1.04%; SBERT-based semantic matching against an index of 14,779 verified hoaxes to catch reworded hoaxes; and source credibility scoring reconciled with the model prediction. Built on a three-tier architecture (React, Express, and Flask) with two-role authentication (user and administrator) using JSON Web Token and bcrypt, it supports text, URL, and image (OCR) analysis, explainable AI, external fact-checker integration, a statistics dashboard, and private history. The system passed 20 black-box test scenarios and 41 automated tests without failures. This research addresses five gaps, namely end-to-end implementation, multi-dataset integration, confidence calibration, semantic matching, and source credibility integration, positioning the system as an accurate, transparent, and trustworthy digital literacy aid.*

Keywords: *hoax detection, IndoBERT, confidence calibration, semantic matching, source credibility, authentication, web system, NLP*

DAFTAR ISI

HALAMAN SAMPUL	0
HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.4.1 Manfaat Akademis	4
1.4.2 Manfaat Praktis	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Teori Utama	7
2.1.1 Hoaks, Misinformasi, dan Disinformasi	7

2.1.2 Pemeriksaan Fakta dan Keterbatasan Verifikasi Manual.....	7
2.1.3 Pemrosesan Bahasa Alami dan Representasi Teks	8
2.1.4 Arsitektur Transformer.....	8
2.1.5 BERT dan Paradigma <i>Pre-training–Fine-tuning</i>	9
2.1.6 IndoBERT	9
2.1.7 Justifikasi Pemilihan Model.....	10
2.2 Teori Pendukung	13
2.2.1 Kalibrasi <i>Confidence</i> dan <i>Temperature Scaling</i>	13
2.2.2 <i>Sentence Embedding</i> dan Kemiripan Semantik	14
2.2.3 Kredibilitas Sumber dan AI yang Layak Dipercaya	14
2.2.4 <i>Explainable AI</i> berbasis <i>Attention</i>	15
2.2.5 Arsitektur Aplikasi Web Tiga Lapis dan REST API	15
2.2.6 Autentikasi, Penyandian Kata Sandi, dan Kontrol Akses Berbasis Peran	16
2.2.7 Pemodelan Sistem dengan UML.....	16
2.3 Penelitian Terdahulu	17
2.4 Gap Penelitian.....	23
2.5 Kerangka Pemikiran.....	24
2.6 Definisi Operasional.....	25
BAB III METODE PENELITIAN.....	27
3.1 Pendekatan Penelitian	27
3.2 Desain Penelitian.....	27
3.3 Objek Penelitian	29
3.4 Instrumen Penelitian.....	30
3.5 Teknik Pengumpulan Data.....	32
3.6 Metode Analisis Data.....	36

3.6.1 Pra-pemrosesan Teks	36
3.6.2 Tokenisasi dan Pembagian Data	38
3.6.3 Konfigurasi <i>Fine-tuning</i>	40
3.6.4 Metrik Evaluasi	42
3.6.5 Kalibrasi <i>Confidence</i> (<i>Temperature Scaling</i>).....	43
3.6.6 Pencocokan Semantik terhadap Korpus Hoaks (<i>Sentence Embedding</i>)	44
3.6.7 Penilaian Kredibilitas Sumber (<i>Rule-Based</i>)	45
3.7 Prosedur Penelitian.....	46
3.7.1 Tahapan Penelitian.....	46
3.7.2 Arsitektur Sistem.....	47
3.7.3 Alur Kerja Fitur Analisis.....	50
3.8 Evaluasi Hasil.....	52
3.8.1 Evaluasi Model.....	52
3.8.2 Pengujian Sistem.....	54
BAB IV PEMBAHASAN.....	57
4.1 Hasil Implementasi Sistem.....	57
4.1.1 Arsitektur dan Komponen yang Terealisasi.....	57
4.1.2 Perancangan Sistem	58
4.1.3 Fitur dan Antarmuka	65
4.1.4 Implementasi Keamanan	78
4.2 Hasil Pelatihan dan Evaluasi Model	79
4.2.1 Proses Pelatihan	79
4.2.2 Hasil Metrik pada Data Uji	82
4.2.3 <i>Confusion Matrix</i> dan Analisis Kesalahan.....	83
4.2.4 Generalisasi: Perbandingan Validasi versus Uji	91

4.2.5 Validasi Eksternal pada Data Nyata.....	91
4.2.6 Perbandingan dengan Pendekatan Penelitian Terdahulu	97
4.2.7 Pembahasan Evaluasi Model.....	100
4.3 Hasil Kalibrasi <i>Confidence</i>	100
4.4 Hasil Fitur Pencocokan Semantik	104
4.4.1 Pemilihan Model <i>Embedding</i> (Studi Ablasi)	104
4.4.2 Hasil Pencocokan pada Contoh Nyata	105
4.5 Hasil Fitur Kredibilitas Sumber	106
4.6 Hasil Pengujian Sistem	110
4.6.1 Pengujian Fungsional (<i>Black-box</i>)	110
4.6.2 Pengujian Otomatis dan <i>Continuous Integration</i>	112
4.6.3 Pengujian Integrasi.....	114
4.7 Pembahasan Umum, Validitas, dan Keterbatasan.....	114
4.7.1 Keterkaitan dengan Celah Penelitian	114
4.7.2 Ancaman terhadap Validitas dan Mitigasinya	115
4.7.3 Keterbatasan	115
4.7.4 Posisi Sistem	117
BAB V KESIMPULAN DAN SARAN.....	118
5.1 Kesimpulan	118
5.1.1 Kesimpulan Akademis	118
5.1.2 Kesimpulan Praktis	120
5.1.3 Kesimpulan Metodologis	121
5.2 Saran untuk Pengembangan Lanjutan.....	121
5.2.1 Pengembangan Model dan Algoritma.....	121
5.2.2 Pengembangan Fitur dan <i>User Experience</i>	123
5.2.3 <i>Deployment</i> dan Infrastruktur	123

5.2.4 Evaluasi dan Validasi Lanjutan.....	124
5.2.5 Kebijakan dan <i>Governance</i>	124
5.3 Rekomendasi Akhir.....	125
DAFTAR PUSTAKA	126
LAMPIRAN.....	131



DAFTAR GAMBAR

Gambar 3.1 Alur tahapan metodologi penelitian.....	29
Gambar 4.1 Diagram <i>use case</i> sistem Checkhoax (dua aktor: <i>User</i> dan <i>Admin</i>).58	
Gambar 4.2 Diagram <i>activity (swimlane)</i> alur login dan pemeriksaan berita.	60
Gambar 4.3 Diagram <i>sequence</i> alur analisis teks.	61
Gambar 4.4 Diagram kelas sistem Checkhoax.	62
Gambar 4.5 <i>Entity Relationship Diagram</i> basis data sistem.	63
Gambar 4.6 <i>Flowchart</i> alur kerja sistem Checkhoax.	64
Gambar 4.7 Halaman Login.	68
Gambar 4.8 Halaman Registrasi.	69
Gambar 4.9 Halaman utama (mode terang).....	70
Gambar 4.10 Halaman Analyser dengan tiga moda masukan (teks, URL, gambar).	71
Gambar 4.11 Moda analisis URL.....	72
Gambar 4.12 Dashboard statistik.....	73
Gambar 4.13 Halaman Riwayat dengan pencarian dan penyaringan.....	74
Gambar 4.14 Mode gelap, halaman Analyser.	75
Gambar 4.15 Mode gelap, Dashboard.	76
Gambar 4.16 Mode gelap, halaman Riwayat.	77
Gambar 4.17 Cuplikan kode pengolahan dataset pada berkas <code>merge_datasets.py</code>	81
Gambar 4.18 <i>Confusion matrix</i> model IndoBERT pada data uji ($n = 4.224$).	84
Gambar 4.19 <i>Confusion matrix</i> model pada data uji eksternal ($n = 200$).	93
Gambar 4.20 <i>Reliability diagram</i> sebelum dan sesudah kalibrasi.	101
Gambar 4.21 Hasil analisis teks: <i>chip</i> “Terkalibrasi” pada tingkat kepercayaan dan kartu “Kemiripan dengan Hoaks Terverifikasi”.	103

Gambar 4.22 Kartu “Penilaian Gabungan (Model + Sumber)” pada analisis URL
dari sumber tepercaya. 109



DAFTAR LAMPIRAN

Lampiran 1. Kartu Asistensi Tugas Akhir	131
Lampiran 2. Curriculum Vitae.....	132
Lampiran 3. Surat Pernyataan & Pengalihan HAKI	133
Lampiran 4. Sertifikat BNSP	135

