

**KLASIFIKASI BERITA HOAKS INDONESIA MENGGUNAKAN
ALGORITMA LIGHTGBM**



LAPORAN MBKM PENELITIAN

**FADHIL DZAKY MUHAMMAD
00000058398**

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

**KLASIFIKASI BERITA HOAKS INDONESIA MENGGUNAKAN
ALGORITMA LIGHTGBM**



LAPORAN MBKM PENELITIAN

FADHIL DZAKY MUHAMMAD

0000058398

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

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

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Fadhil Dzaky Muhammad
Nomor Induk Mahasiswa : 00000058398
Program Studi : Informatika

Skripsi dengan judul:

Klasifikasi Berita Hoaks Indonesia Menggunakan Algoritma LightGBM

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, 03 Januari 2025



(Fadhil Dzaky Muhammad)

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

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Fadhil Dzaky Muhammad
NIM : 00000058398
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Klasifikasi Berita Hoaks Indonesia
Menggunakan Algoritma LightGBM

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, 03 Januari 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

Fadhil Dzaky Muhammad

**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

”Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Ia mendapat pahala (dari kebaikan) yang diusahakannya dan ia mendapat siksa (dari kejahatan) yang dikerjakannya.”

QS. Al-Baqarah:286

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat-Nya sehingga penelitian ini dapat diselesaikan. Penelitian ini bertujuan untuk memberikan kontribusi akademis dalam upaya mendeteksi dan mengurangi penyebaran berita hoaks di Indonesia.

Dalam penyelesaian penelitian ini, ucapan terima kasih disampaikan 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 Assoc. Prof. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya penelitian ini.
5. Keluarga atas dukungan material dan moral sehingga penelitian ini dapat diselesaikan.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pengembangan metode deteksi berita hoaks di Indonesia.

Tangerang, 03 Januari 2025

UNIVERSITAS
MULTIMEDIA
NUSANTARA


Fadhil Dzaky Muhammad

KLASIFIKASI BERITA HOAKS INDONESIA MENGGUNAKAN ALGORITMA LIGHTGBM

Fadhil Dzaky Muhammad

ABSTRAK

Berita hoaks merupakan ancaman serius terhadap demokrasi, stabilitas sosial, dan integritas politik di Indonesia, terutama dalam konteks pemilu dan isu-isu politik lainnya. Penyebaran berita hoaks dapat memengaruhi opini publik secara negatif, memperburuk polarisasi masyarakat, dan memicu konflik sosial. Penelitian ini bertujuan untuk mengembangkan model klasifikasi berita hoaks yang andal dan efisien menggunakan algoritma LightGBM. Dataset yang digunakan mencakup 59.850 entri berita yang telah melalui proses pembersihan data dan preprocessing teks. Untuk ekstraksi fitur, metode Term Frequency-Inverse Document Frequency (TF-IDF) diterapkan guna mengonversi data teks ke dalam format numerik yang dapat diproses oleh model machine learning. Parameter model dioptimalkan menggunakan Optuna, sebuah framework hyperparameter tuning berbasis Bayesian Optimization. Hasil penelitian menunjukkan bahwa LightGBM mampu mencapai akurasi yang sangat tinggi sebesar 99,23% dengan efisiensi pemrosesan yang baik.

Kata kunci: Berita hoaks, Indonesia, Klasifikasi, LightGBM, Machine learning,



**CLASSIFICATION OF INDONESIAN HOAX NEWS USING LIGHTGBM
ALGORITHM**

Fadhil Dzaky Muhammad

ABSTRACT

Fake news poses a serious threat to democracy, social stability, and political integrity in Indonesia, particularly in the context of elections and other political issues. The spread of fake news can negatively influence public opinion, worsen societal polarization, trigger social conflicts, and impact decision-making in electoral processes. This study aims to develop a reliable and efficient fake news classification model using the LightGBM algorithm. The dataset used consists of 59,850 news entries, with fake news sourced from MAFINDO and factual news collected from reputable news portals such as KOMPAS, TEMPO, and Detik. The dataset underwent data cleaning and text preprocessing. For feature extraction, the Term Frequency-Inverse Document Frequency (TF-IDF) method was applied to convert textual data into numerical formats suitable for machine learning models. Model parameters were optimized using Optuna, a hyperparameter tuning framework based on Bayesian Optimization. The results show that LightGBM achieves an exceptionally high accuracy of 99.23% with excellent processing efficiency.

Keywords: *Classification, Hoax news, Indonesia, LightGBM, Machine learning*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iii
HALAMAN PERSEMBAHAN/MOTO	iv
KATA PENGANTAR	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR KODE	xii
DAFTAR LAMPIRAN	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Urgensi Penelitian	3
1.5 Luaran Penelitian	4
1.6 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Berita Hoaks	5
2.2 Machine Learning	5
2.3 Text Pre-processing	6
2.3.1 Normalisasi Teks	7
2.3.2 Tokenisasi	8
2.3.3 Penghapusan Stop-word	8
2.3.4 Stemming	8
2.4 Term Frequency-Inverse Document Frequency (TF-IDF)	9
2.5 Decision Tree	10
2.6 <i>Ensemble Learning</i>	11
2.6.1 <i>Bagging</i>	12
2.6.2 <i>Boosting</i>	13
2.7 <i>Gradient Boosting Decision Tree</i>	15
2.8 Light Gradient Boosting Machine (LightGBM)	17
2.8.1 <i>Histogram Based Algorithm</i>	18
2.8.2 <i>Gradient-based One-Side Sampling (GOSS)</i>	18
2.8.3 <i>Exclusive Feature Bundling (EFB)</i>	20
BAB 3 METODOLOGI PENELITIAN	23
3.1 Metodologi Penelitian	23
3.2 Studi Literatur	24
3.3 Pengumpulan Data	25
3.4 Preprocessing Data	27
3.5 Pembagian Dataset dan Ekstraksi Fitur	30
3.6 Perancangan Model LightGBM	31
3.6.1 Inisialisasi Model	33
3.6.2 Optimasi Parameter	33
3.6.3 Pelatihan Model	34
3.7 Evaluasi Model	34

3.8	Gap Penelitian	37
BAB 4	HASIL DAN PEMBAHASAN	39
4.1	Dataset	39
4.2	Preprocessing	40
4.3	Pembagian Dataset dan Ekstraksi Fitur	42
4.4	Optimasi dan Penyesuaian Parameter	45
4.5	Evaluasi Model	47
4.5.1	Analisis Performa Model	47
4.5.2	Hasil Cross-Validation	49
BAB 5	SIMPULAN DAN SARAN	52
5.1	Simpulan	52
5.2	Saran	52
DAFTAR PUSTAKA	54



DAFTAR TABEL

Tabel 3.1	Spesifikasi Perangkat Keras dan Perangkat Lunak Penelitian	23
Tabel 3.2	Sumber dan Karakteristik Pengumpulan Data	26
Tabel 3.3	Daftar <i>Stop Words</i> yang Digunakan	28
Tabel 3.4	Ruang Parameter untuk Optimasi LightGBM	33
Tabel 3.5	Perbandingan Penelitian Terdahulu	37
Tabel 4.1	Original Text Statistics	40
Tabel 4.2	Text Preprocessing Steps	41
Tabel 4.3	Preprocessed Text Statistics	41
Tabel 4.4	Distribusi Dataset	43
Tabel 4.5	Contoh Perhitungan TF-IDF	43
Tabel 4.6	Statistik Fitur TF-IDF	44
Tabel 4.7	Hasil Tuning Parameter Awal LightGBM	45
Tabel 4.8	Parameter Hasil Penyesuaian Model LightGBM	46
Tabel 4.9	Performance Metrics for Training and Testing	48
Tabel 4.10	Detail Metrik LightGBM 10-Fold Cross-Validation	49
Tabel 4.11	Metrik Agregasi LightGBM 10-Fold Cross-Validation	50



DAFTAR GAMBAR

Gambar 2.1	<i>Preprocessing</i>	7
Gambar 3.1	<i>Flowchart Penelitian</i>	24
Gambar 3.2	<i>Pengumpulan Data</i>	25
Gambar 3.3	<i>Preprocessing</i>	27
Gambar 3.4	<i>Dataset Split & TF-IDF Vectorization</i>	30
Gambar 3.5	<i>Tahapan Perancangan Model LightGBM</i>	32
Gambar 4.1	<i>Distribusi Label</i>	39
Gambar 4.2	<i>Jumlah Kata di Original Dataset</i>	40
Gambar 4.3	<i>Jumlah Kata di Preprocessed Dataset</i>	42
Gambar 4.4	<i>Top 20 Mean TF-IDF Score</i>	44
Gambar 4.5	<i>Analisis Parameter Importance LightGBM</i>	46
Gambar 4.6	<i>Confusion Matrix 10-Fold Cross-Validation</i>	47
Gambar 4.7	<i>Akurasi dari 10-Fold Cross-Validation</i>	49
Gambar 4.8	<i>Confusion Matrix 10-Fold Cross-Validation</i>	50



DAFTAR KODE



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	59
Lampiran 2	Formulir Bimbingan	71
Lampiran 3	Cover Letter	72
Lampiran 4	MBKM Research Card	73
Lampiran 5	Lembar Verifikasi	74
Lampiran 6	Daily Task	75
Lampiran 7	Surat Kerjasama UNIIC dengan LPPM	89
Lampiran 8	Draft Paper	90

