

TUGAS AKHIR

**PERBANDINGAN KINERJA *XGBOOST* DAN *LIGHTGBM*
DENGAN *HYPERPARAMETER TUNING* PADA KLASIFIKASI
BERITA *HOAX* INDONESIA**

Disusun untuk Melengkapi Tugas dan Memenuhi Syarat Kelulusan
Program Strata 1 Jurusan Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember



Wildan Ayu Shafira
2210651087

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2026

PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini:

Nama : Wildan Ayu Shafira
NIM : 2210651087
Program Studi : Teknik Informatika
Judul Tugas Akhir : Perbandingan Kinerja *XGBoost* dan *LightGBM* dengan *Hyperparameter Tuning* pada Klasifikasi Berita *Hoax* Indonesia

Dengan ini menyatakan bahwa karya ilmiah yang saya susun adalah murni hasil pemikiran, penelitian, dan penyusunan saya sendiri. Segala sumber informasi yang berasal atau dikutip dari karya orang lain, baik berupa publikasi, artikel, buku, maupun sumber lainnya, telah disebutkan dengan jelas di dalam daftar pustaka sesuai dengan kaidah penulisan ilmiah.

Saya menyadari bahwa dalam proses penyusunan Tugas Akhir ini, saya memanfaatkan bantuan teknologi kecerdasan buatan (AI) seperti *ChatGPT/Grammarly/Alat AI* lain hanya sebatas untuk:

1. Membantu dalam penyusunan tata bahasa, parafrasa, atau perbaikan redaksi.
2. Memberikan inspirasi ide awal atau kerangka berpikir, yang selanjutnya saya kembangkan secara mandiri.
3. Membantu dalam pengecekan konsistensi format, ejaan, dan tata tulis.

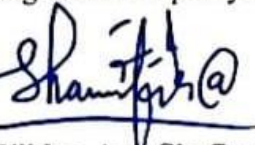
Saya menegaskan bahwa tidak ada bagian dari karya ilmiah ini yang seluruhnya dibuat oleh AI tanpa keterlibatan pemikiran kritis saya sendiri. Tanggung jawab penuh atas isi, keaslian, dan kebenaran karya ilmiah ini ada pada diri saya.

Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima segala konsekuensi sesuai dengan peraturan yang berlaku di Universitas Muhammadiyah Jember.

Jember, 29 Juli 2026

Yang membuat pernyataan




(Wildan Ayu Shafira)

LEMBAR PERSETUJUAN

PERBANDINGAN KINERJA *XGBOOST* DAN *LIGHTGBM* DENGAN *HYPERPARAMETER TUNING* PADA KLASIFIKASI BERITA *HOAX* INDONESIA

Diajukan Oleh:

Wildan Ayu Shafira

2210651087

Telah disetujui bahwa laporan Tugas Akhir ini untuk diajukan pada
Sidang Tugas Akhir sebagai salah satu syarat kelulusan dan
mendapatkan gelar Sarjana Komputer (S.Kom)

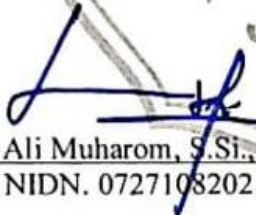
di

Universitas Muhammadiyah Jember

Menyetujui,

Pembimbing 1

Pembimbing 2


Lutfi Ali Muharom, S.Si., M.Si
NIDN. 0727108202


Daryanto, S.Kom., M.Kom
NIDN. 0707077203

LEMBAR PENGESAHAN

**PERBANDINGAN KINERJA XGBOOST DAN LIGHTGBM DENGAN
HYPERPARAMETER TUNING PADA KLASIFIKASI BERITA HOAX
INDONESIA**

Oleh:

Wildan Ayu Shafira

2210651087

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 29 Bulan 07 Tahun 2026 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Dosen Penguji:
Penguji I


Zainul Arifin, S.Si., M.Kom
NIDN. 0714078102

Dosen Pembimbing:
Pembimbing I


Lutfi Ali Muharom, S.Si., M.Si
NIDN. 0727108202

Penguji II


Henny Wahyu Sulisty, S.Kom, M.Kom
NIDN. 0718088309

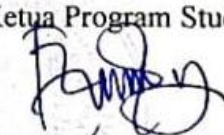
Pembimbing II


Daryanto, S.Kom., M.Kom
NIDN. 0707077203

Mengesahkan,
Dekan Fakultas Teknik


Prof. Dr. Ir. Muhlar, S.T., M.T., IPM
NIDN. 0010067301

Mengetahui,
Ketua Program Studi Teknik Inform


Rosita Yanuarti, S.Kom., M.Cs
NIDN. 0629018601

KATA PENGANTAR

Alhamdulillah rabbil'alamin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta hidayah-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Perbandingan Kinerja *XGBoost* dan *LightGBM* dengan *Hyperparameter Tuning* pada Klasifikasi Berita *Hoax* Berbahasa Indonesia” dengan baik. Shalawat serta salam senantiasa tercurah kepada Nabi Muhammad SAW yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh dengan ilmu pengetahuan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember. Selama proses penyusunan tugas akhir ini, penulis menyadari bahwa penyelesaian penelitian ini tidak terlepas dari doa, bimbingan, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa hormat dan ucapan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, sehingga tugas akhir ini dapat terselesaikan.

Akhir kata, penulis berharap semoga tugas akhir ini dapat memberikan manfaat, menambah wawasan, serta menjadi salah satu referensi bagi pembaca maupun peneliti selanjutnya, khususnya dalam bidang machine learning dan klasifikasi teks berbahasa Indonesia.

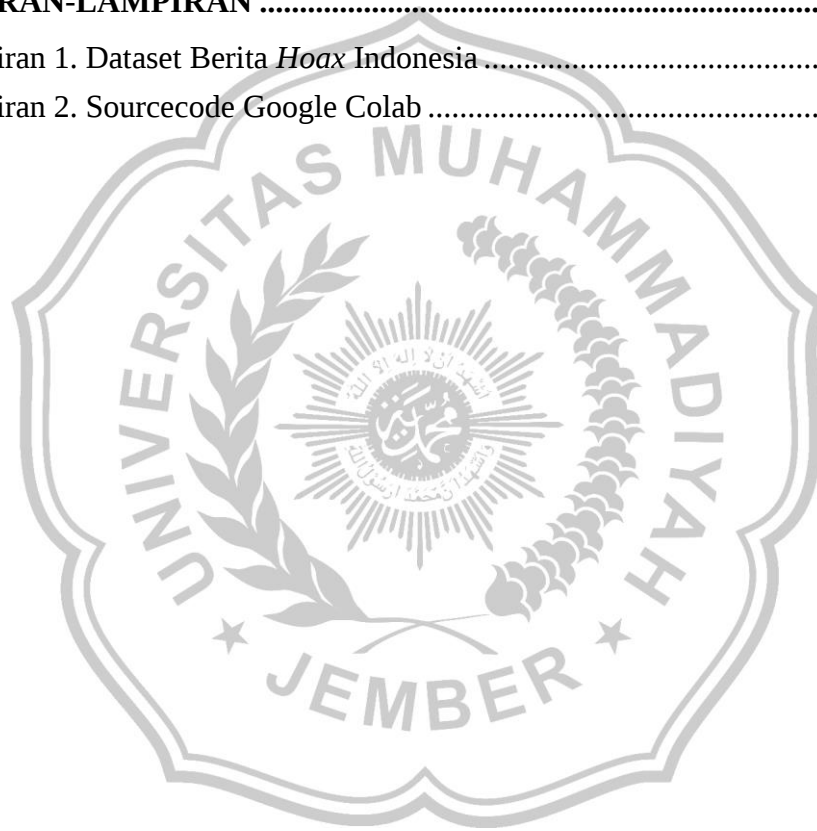
DAFTAR ISI

PENYATAAN ORISINALITAS	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
MOTTO	iv
PERSEMBAHAN.....	v
ABSTRAK	viii
ABSTRACT	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian	6
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu	7
2.2 Berita <i>Hoax</i>	8
2.2.1 Pengertian Berita <i>Hoax</i>	8
2.2.2 Karakteristik Berita <i>Hoax</i>	8
2.2.3 Dampak Penyebaran <i>Hoax</i>	9
2.2.4 Pentingnya Deteksi <i>Hoax</i> Menggunakan Teknologi.....	9
2.3 <i>Data Mining</i>	10
2.3.1 Pengertian <i>Data Mining</i>	10
2.3.2 Fungsi <i>Data Mining</i>	10
2.3.3 Tahapan <i>Data Mining</i>	10
2.4 <i>Machine Learning</i>	11
2.4.1 Pengertian <i>Machine Learning</i>	11
2.4.2 Jenis-Jenis <i>Machine Learning</i>	11
2.5 Klasifikasi Teks.....	12

2.5.1 Pengertian Klasifikasi Teks.....	12
2.5.2 Proses Klasifikasi Teks	12
2.5.3 Penerapan Klasifikasi Teks dalam Deteksi Berita <i>Hoax</i>	13
2.6 <i>Text Preprocessing</i>	13
2.6.1 Pengertian <i>Text Preprocessing</i>	13
2.6.2 Tahapan <i>Text Preprocessing</i>	14
2.6.3 Tujuan <i>Preprocessing</i> dalam <i>Machine Learning</i>	15
2.7 TF-IDF (<i>Term Frequency - Inverse Document Frequency</i>)	15
2.7.1 Pengertian TF-IDF	15
2.7.2 Cara Kerja TF-IDF	15
2.7.3 Keunggulan TF-IDF dalam Klasifikasi Teks.....	16
2.8 Algoritma <i>XGBoost</i>	16
2.8.1 Pengertian <i>XGBoost</i>	16
2.8.2 Konsep Dasar <i>Boosting</i>	17
2.8.3 Cara Kerja <i>XGBoost</i>	17
2.8.4 Kelebihan dan Kekurangan <i>XGBoost</i>	18
2.9 Algoritma <i>LightGBM</i>	18
2.9.1 Pengertian <i>LightGBM</i>	18
2.9.2 Konsep Gradient Boosting dalam <i>LightGBM</i>	18
2.9.3 Cara Kerja <i>LightGBM</i>	19
2.9.4 Keunggulan <i>LightGBM</i> dibanding Algoritma Lain	19
2.10 <i>Hyperparameter tuning</i>	20
2.10.1 Pengertian <i>Hyperparameter</i>	20
2.10.2 Perbedaan Parameter dan <i>Hyperparameter</i>	21
2.10.3 Tujuan <i>Hyperparameter tuning</i>	21
2.10.4 Metode <i>Hyperparameter tuning</i>	21
2.10.5 Pengaruh <i>Tuning</i> terhadap Performa Model	22
2.11 Evaluasi Model Klasifikasi	22
2.11.1 Konsep Evaluasi Model	22
2.11.2 <i>Confusion Matrix</i>	23
2.11.3 Metrik Evaluasi	23
BAB 3 METODOLOGI PENELITIAN	25
3.1 Alur Penelitian	25
3.2 Jenis dan Pendekatan Penelitian.....	25
3.3 Bahan Penelitian.....	26

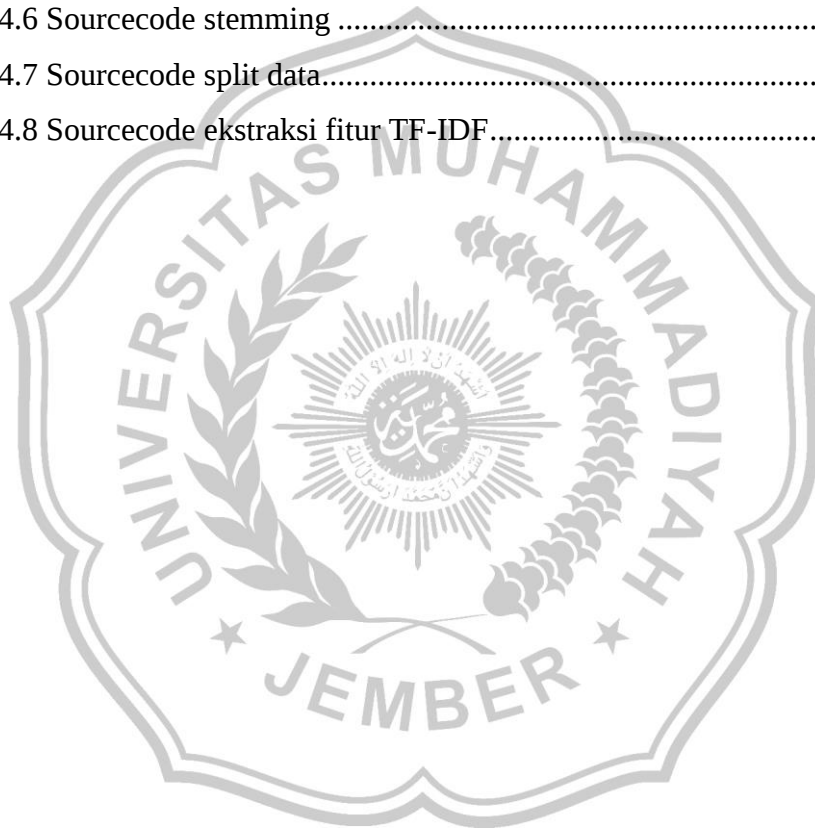
3.4 Peralatan Penelitian.....	26
3.4.1 Perangkat Keras	26
3.4.2 Perangkat Lunak.....	27
3.4.3 <i>Library</i> dan <i>Tools</i>	27
3.5 Metode Penelitian.....	27
3.5.1 <i>Business Understanding</i>	28
3.5.2 <i>Data Understanding</i>	28
3.5.3 <i>Data Preparation</i>	29
3.5.4 <i>Modeling</i>	35
3.5.5 <i>Evaluation</i>	36
3.5.6 <i>Deployment</i>	39
3.6 <i>Hyperparameter tuning</i>	39
3.6.1 Metode <i>Hyperparameter tuning</i>	39
3.6.2 Parameter pada <i>XGBoost</i>	40
3.6.3 Parameter pada <i>LightGBM</i>	40
3.6.4 Prosedur <i>Hyperparameter tuning</i>	41
3.7 Skenario Pengujian.....	41
3.7.1 Skenario <i>Baseline</i> (Tanpa <i>Tuning</i>).....	42
3.7.2 Skenario Setelah <i>Hyperparameter tuning</i>	42
3.7.3 <i>Cross-Validation</i> (k=5).....	42
3.7.4 Metrik Evaluasi	42
3.8 Variabel Penelitian.....	43
BAB 4 HASIL DAN PEMBAHASAN.....	45
4.1 Pengumpulan Data	45
4.2 Data Preparation.....	45
4.2.1 Hasil <i>Preprocessing</i> Teks	45
4.2.2 Hasil Pembagian Data	48
4.2.3 Hasil Pembobotan TF-IDF	49
4.3 Hasil Pemodelan Baseline.....	50
4.3.1 <i>XGBoost</i> Baseline.....	50
4.3.2 <i>LightGBM</i> Baseline	51
4.3.3 Perbandingan Baseline	52
4.4 Hasil <i>Hyperparameter tuning</i>	53
4.4.1 <i>Grid Search XGBoost</i>	53
4.4.2 <i>Grid Search LightGBM</i>	53

4.5 Hasil Evaluasi dan Perbandingan Model	54
4.6 Pembahasan.....	55
4.6.1 Analisis Pengaruh <i>Hyperparameter tuning</i> terhadap kinerja Model ...	55
4.6.2 Perbandingan dengan Penelitian Terdahulu.....	56
4.7 Model Terbaik Hasil Penelitian	56
BAB 5 KESIMPULAN & SARAN	57
5.1 Kesimpulan	57
5.2 Saran.....	58
DAFTAR PUSTAKA.....	59
LAMPIRAN-LAMPIRAN	63
Lampiran 1. Dataset Berita <i>Hoax</i> Indonesia	63
Lampiran 2. Sourcecode Google Colab	64



DAFTAR GAMBAR

Gambar 2.1 Tahapan CRISP-DM (Sumber: Chapman <i>et al.</i> ,2000).....	11
Gambar 3.1 Diagram alur penelitian.....	25
Gambar 4.1 Sourcecode cleaning data.....	45
Gambar 4.2 Sourcecode hapus data duplikat dan data kosong.....	46
Gambar 4.3 Sourcecode case folding.....	46
Gambar 4.4 Sourcecode tokenization	46
Gambar 4.5 Sourcecode stopwords removal	47
Gambar 4.6 Sourcecode stemming	47
Gambar 4.7 Sourcecode split data.....	48
Gambar 4.8 Sourcecode ekstraksi fitur TF-IDF.....	49



DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	7
Tabel 2.2 <i>Confusion matrix</i>	23
Tabel 3.1 Distribusi kelas dataset	28
Tabel 3.2 Contoh data awal.....	29
Tabel 3.3 Hasil data cleaning	30
Tabel 3.4 Hasil <i>case folding</i>	30
Tabel 3.5 Hasil tokenisasi	31
Tabel 3.6 Hasil <i>stopward removal</i>	31
Tabel 3.7 Hasil <i>stemming</i>	32
Tabel 3.8 Nilai TF Dokumen D1 (Non- <i>Hoax</i>).....	33
Tabel 3.9 Perhitungan Nilai IDF (N = 4 Dokumen)	33
Tabel 3.10 Matriks Nilai TF-IDF.....	34
Tabel 3.11 Contoh hasil confusion matrix	37
Tabel 3.12 Parameter <i>XGBoost</i>	40
Tabel 3.13 Parameter <i>LightGBM</i>	40
Tabel 3.14 Variabel penelitian	44
Tabel 4.1 Hasil preprocessing.....	48
Tabel 4.2 Hasil ekstraksi fitur TF IDF	49
Tabel 4.3 Classification report <i>XGBoost</i> baseline.....	50
Tabel 4.4 Confusion matrix <i>XGBoost</i> baseline.....	51
Tabel 4.5 Confusion matrix <i>LightGBM</i> baseline	51
Tabel 4.6 Classification Report <i>LightGBM</i> Baseline.....	52
Tabel 4.7 Perbandingan hasil evaluasi model baseline.....	52
Tabel 4.8 Kombinasi parameter terbaik <i>Grid Search XGBoost</i>	53
Tabel 4.9 Kombinasi parameter terbaik <i>Grid Search LightGBM</i>	54
Tabel 4.10 Perbandingan keseluruhan kinerja model	54