

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
XGBOOST DALAM MEMPREDIKSI TIPE KECELAKAAN
LALU LINTAS**



SKRIPSI

**BHAKTA WACHELA NAIDU
00000041722**

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

**PERBANDINGAN ALGORITMA RANDOM FOREST DAN
XGBOOST DALAM MEMPREDIKSI TIPE KECELAKAAN
LALU LINTAS**



SKRIPSI

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

BHAKTA WACHELA NAIDU

00000041722

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Bhakta Wachela Naidu

Nomor Induk Mahasiswa : 00000041722

Program Studi : Informatika

Skripsi dengan judul:

Perbandingan Algoritma Random Forest dan XGBoost dalam Memprediksi Tipe Kecelakaan Lalu Lintas

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, 1 Juli 2025



(Bhakta Wachela Naidu)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MEMPREDIKSI TIPE KECELAKAAN LALU LINTAS

oleh

Nama : Bhakta Wachela Naidu
NIM : 00000041722
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

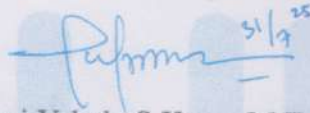
Telah diujikan pada hari Rabu, 23 Juli 2025

Pukul 10.00 s/s 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

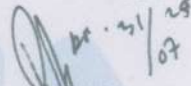
Ketua Sidang



(Sy Yuliani Yakub, S.Kom., M.T. PhD)

NIDN: 0411037904

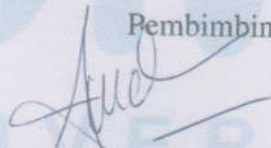
Penguji



(Angga Aditya Permana, S.Kom.,
M.Kom.)

NIDN: 0407128901

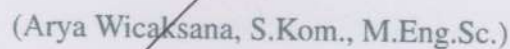
Pembimbing



(Anak Agung Ngurah Ananda Kusuma, B.Eng., M.Eng., Ph.D.)

NIDK: 08984101024

Ketua Program Studi Informatika,



(Arya Wicaksana, S.Kom., M.Eng.Sc.)

NIDN: 0315109103

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

Yang bertanda tangan di bawah ini:

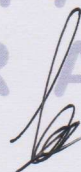
Nama : Bhakta Wachela Naidu
NIM : 00000041722
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Perbandingan Algoritma Random
Forest dan XGBoost dalam
Memprediksi Tipe Kecelakaan
Lalu Lintas

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, 1 Juli 2025

Yang menyatakan



Bhakta Wachela Naidu

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



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSEMBAHAN / MOTTO

"You can change what you do, but you can't change what you want."

Thomas Shelby



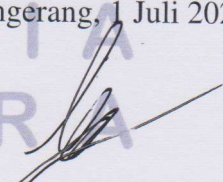
KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Perbandingan Algoritma Random Forest dan XGBoost dalam Memprediksi Tipe Kecelakaan Lalu Lintas 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 skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi 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., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Anak Agung Ngurah Ananda Kusuma, B.Eng., M.Eng., Ph.D., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 1 Juli 2025


Bhakta Wachela Naidu

PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MEMPREDIKSI TIPE KECELAKAAN LALU LINTAS

Bhakta Wachela Naidu

ABSTRAK

Kecelakaan lalu lintas merupakan salah satu permasalahan utama dalam keselamatan transportasi yang terus meningkat setiap tahunnya. Sebagian besar penelitian terdahulu berfokus pada faktor manusia seperti usia, jenis kelamin, dan pengaruh zat psikoaktif, padahal faktor lingkungan seperti kondisi jalan, cuaca, serta infrastruktur juga memiliki pengaruh signifikan terhadap penyebab kecelakaan lalu lintas. Penelitian ini bertujuan untuk memprediksi kategori penyebab utama kecelakaan yang dibedakan menjadi *human error* dan *non-human error*, dengan mempertimbangkan faktor-faktor lingkungan menggunakan algoritma *random forest* dan *XGBoost*. Dilakukan pembangunan model awal tanpa *parameter tuning*. Kemudian, dilakukan *hyperparameter tuning* menggunakan *GridSearchCV* untuk meningkatkan performa model. Model dievaluasi menggunakan *confusion matrix* dan *classification report*. Hasil pengujian menunjukkan bahwa algoritma *XGBoost* memiliki akurasi sebesar 92,49% dan *random forest* sebesar 92,5%. Analisis *feature importance* memperlihatkan perbedaan preferensi fitur dari kedua model, yang mencerminkan pendekatan internal masing-masing algoritma dalam membentuk keputusan. Berdasarkan hasil evaluasi, dapat disimpulkan bahwa kedua model sama-sama mampu memprediksi tipe kecelakaan berbasis lingkungan dengan performa yang sebanding.

Kata kunci: Classification report, Hyperparameter Tuning, Non-Human Error, Random forest, XGBoost

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

Comparison of Random Forest and XGboost Algorithms in Predicting Traffic Accident Types

Bhakta Wachela Naidu

ABSTRACT

Traffic accidents are one of the main issues in transportation safety, with cases continuing to rise each year. Most previous studies have focused on human-related factors such as age, gender, and the influence of psychoactive substances, whereas environmental factors such as road conditions, weather, and infrastructure also have a significant impact on the causes of traffic accidents. This study aims to predict the primary contributory cause category of accidents—classified as human error or non-human error—by considering environmental factors using the Random forest and XGBoost algorithm. An initial model was built without parameter tuning. Then, hyperparameter tuning was performed using GridSearchCV to improve model performance. The models were evaluated using confusion matrix and classification report. The test results showed that the XGBoost algorithm achieved an accuracy of 92,49%, while random forest reached 92,5%. The feature importance analysis revealed different feature preferences between the two models, reflecting each algorithm's internal approach to decision-making. Based on the evaluation results, it can be concluded that both models are capable of predicting accident types based on environmental factors with comparable performance.

Keywords : *Classification Report, Hyperparameter Tuning, Non-Human Error, Random Forest, XGBoost*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR KODE	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 <i>Machine Learning</i>	6
2.2 <i>Decision Trees</i>	6
2.3 <i>Ensemble Learning</i>	7
2.4 <i>Random Forest</i>	8
2.5 <i>XGBoost</i>	10
2.6 <i>Confusion Matrix</i>	10
BAB 3 METODOLOGI PENELITIAN	12
3.1 Metodologi Penelitian	12
3.2 Alur penelitian	12
3.3 Studi Literatur	13
3.4 Pengumpulan data	13
3.5 Pre-processing Data	15
3.5.1 Handling Missing Values	15
3.5.2 Duplicate Detection and Removal	15
3.5.3 Removal of Unusable Label	15
3.5.4 Feature Filtering	16
3.5.5 Categorization of Primary Contributory Cause	16
3.5.6 Encoding Categorical Data	18
3.6 Pembangunan Model Awal	18
3.7 <i>Hyperparameter Tuning</i> dan Iterasi Model Kedua	18
3.8 Evaluasi	19
BAB 4 HASIL DAN DISKUSI	20
4.1 Spesifikasi Sistem	20
4.2 Pengumpulan Data	20
4.2.1 Import Libraries	20
4.2.2 Import Data	21
4.3 Pre-processing Data	21

4.3.1	Handling Missing Values	21
4.3.2	Duplicate Detection and Removal	22
4.3.3	Removal of Unusable Label	23
4.3.4	Feature Filtering	24
4.3.5	Categorization of Primary Contributory Cause	24
4.3.6	Encoding Data Kategorikal	27
4.4	Pembangunan Model	29
4.4.1	Pembagian Data untuk Pelatihan dan Pengujian	29
4.4.2	Pembangunan Model Awal	30
4.4.3	<i>Hyperparameter tuning</i> dan Iterasi Model Kedua	32
4.5	Evaluasi	35
4.5.1	Evaluasi Akurasi pada Data Latih dan Data Uji	35
4.5.2	Analisis <i>Confusion Matrix dan Classification Report</i>	36
4.5.3	Analisis <i>Feature Importance</i>	40
BAB 5	SIMPULAN DAN SARAN	50
5.1	Simpulan	50
5.2	Saran	50
DAFTAR PUSTAKA		51



DAFTAR TABEL

Tabel 2.1	Confusion Matrix	10
Tabel 3.1	Studi literatur	13
Tabel 3.2	Kategorisasi awal	16
Tabel 4.1	Hasil kategorisasi	27
Tabel 4.2	<i>Confusion matrix</i> model RF awal	31
Tabel 4.3	<i>Confusion matrix</i> model XGB awal	31
Tabel 4.4	Perbandingan akurasi model	34
Tabel 4.5	<i>Classification Report</i> Model RF	39
Tabel 4.6	<i>Classification Report</i> Model XGB	40
Tabel 4.7	Perbandingan persentase <i>feature importance</i> kedua model	42
Tabel 4.8	<i>Classification Report</i> Model <i>Random Forest</i> setelah feature selection	48
Tabel 4.9	<i>Classification Report</i> Model <i>XGBoost</i> setelah feature selection	48
Tabel 4.10	Perbandingan performa model setelah feature selection	49
Tabel 4.11	Perbandingan waktu komputasi setelah feature selection	49



DAFTAR GAMBAR

Gambar 2.1	Tipe-tipe algoritma ML	6
Gambar 2.2	Visualisasi DT	7
Gambar 2.3	Metode bagging	8
Gambar 2.4	Metode boosting	8
Gambar 3.1	Diagram alur penelitian	12
Gambar 4.1	Tahapan <i>preprocess</i> data	21
Gambar 4.2	Tahapan <i>handling missing values</i>	22
Gambar 4.3	Hasil implementasi untuk menemukan data duplikasi	23
Gambar 4.4	Penghapusan label " <i>UNABLE TO DETERMINE</i> "	23
Gambar 4.5	Hasil <i>encoding</i>	28
Gambar 4.6	Visualisasi <i>confusion matrix</i> Model RF	37
Gambar 4.7	Visualisasi <i>confusion matrix</i> Model XGB	38
Gambar 4.8	Visualisasi <i>feature importance</i> Model RF	43
Gambar 4.9	Visualisasi <i>feature importance</i> model XGB	44



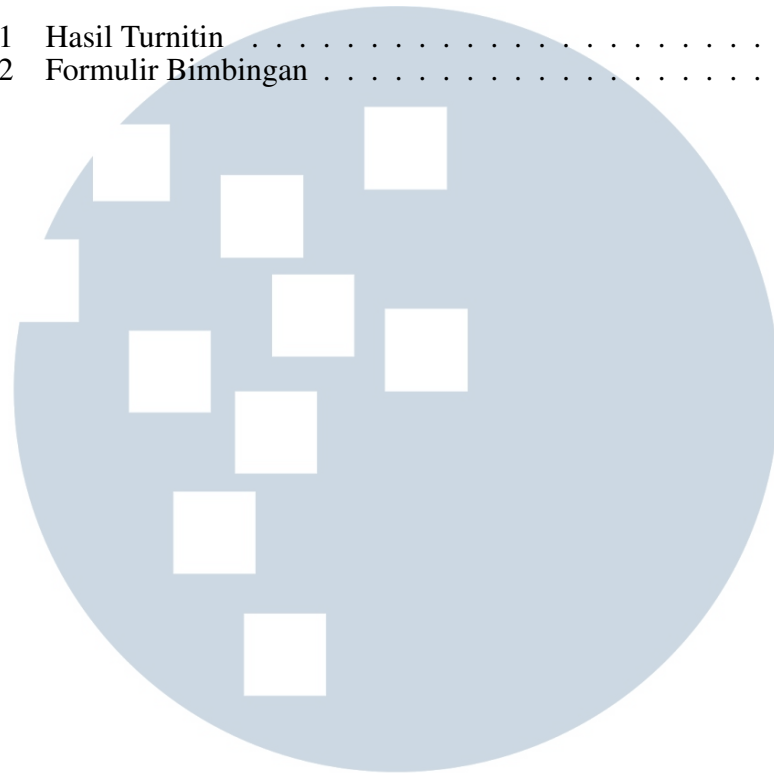
DAFTAR KODE

Kode 4.1	<i>Import libraries</i> yang digunakan pada penelitian	20
Kode 4.2	<i>Import</i> dan membaca dataset	21
Kode 4.3	Cek duplikasi	22
Kode 4.4	<i>Drop duplicate value</i>	23
Kode 4.5	Implementasi <i>feature filtering</i>	24
Kode 4.6	Implementasi kategorisasi <i>human error</i> dan <i>non-human error</i> . .	24
Kode 4.7	Implementasi kategorisasi	27
Kode 4.8	Implementasi <i>encoder</i>	28
Kode 4.9	Implementasi kategorisasi	29
Kode 4.10	Pembangunan model <i>random forest</i>	30
Kode 4.11	Pembangunan model <i>XGBoost</i>	30
Kode 4.12	Pembangunan model <i>XGBoost</i> menggunakan <i>gridsearch</i> untuk <i>hyperparameter tuning</i>	32
Kode 4.13	Pembangunan model <i>XGBoost</i> menggunakan <i>gridsearch</i> untuk <i>hyperparameter tuning</i>	33
Kode 4.14	Perbandingan akurasi latih dan akurasi <i>testing</i>	35
Kode 4.15	Implementasi <i>heatmap confusion matrix</i>	36
Kode 4.16	Perbandingan <i>Feature Importance</i> dari kedua model	41
Kode 4.17	Visualisasi <i>top 10 feature importance</i>	42
Kode 4.18	Pengambilan 7 fitur teratas	45
Kode 4.19	<i>Test split</i> setelah pemilihan 7 fitur teratas	46
Kode 4.20	<i>Retraining</i> model RF setelah pemilihan 7 fitur teratas	46
Kode 4.21	<i>Retraining</i> model XGB setelah pemilihan 7 fitur teratas	47



DAFTAR LAMPIRAN

Lampiran 1	Hasil Turnitin	54
Lampiran 2	Formulir Bimbingan	63



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA