

TUGAS AKHIR

**PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI
MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI
*ENSEMBLE***



DISUSUN OLEH:

IBNU RAJU HUMAM

NIM. 21106050047

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2025

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1677/Un.02/DST/PP.00.9/08/2025

Tugas Akhir dengan judul : **PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI ENSEMBLE**

yang dipersiapkan dan disusun oleh:

Nama : **IBNU RAJU HUMAM**
Nomor Induk Mahasiswa : **21106050047**
Telah diujikan pada : **Senin, 21 Juli 2025**
Nilai ujian Tugas Akhir : **A-**

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Prof. Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom., IPM., ASEAN Eng.
SIGNED

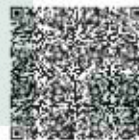
Valid ID: 6892b56922999



Penguji I

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 688ae5e6e132



Penguji II

Dr. Fitri Wulandari, S.Si., M.Kom.
SIGNED

Valid ID: 688c4e21937d7



Yogyakarta, 21 Juli 2025
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 689af2d05b50b

PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Ibnu Raju Humam
NIM : 21106050047
Program Studi : Informatika
Fakultas : Sains dan Teknologi
Universitas : UIN Sunan Kalijaga Yogyakarta

menyatakan dengan sesungguhnya, bahwa skripsi saya yang berjudul “PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI *ENSEMBLE*” merupakan penelitian saya sendiri, tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Karya ini diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Program Studi Informatika pada Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

Yogyakarta, 16 Juli 2025

Penulis



Ibnu Raju Humam

NIM. 21106050047

SURAT PERSETUJUAN TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi/Tugas Akhir
Lampiran : -

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
Di Yogyakarta

Assalammu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Ibnu Raju Humam
NIM : 21106050047
Judul Skripsi : PENGARUH KARAKTERISTIK DATA DAN
OPTIMALISASI MODEL TERHADAP KINERJA
ALGORITMA KLASIFIKASI *ENSEMBLE*

sudah dapat diajukan kembali kepada Program Studi Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Informatika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara dapat segera di-munagasyah-kan. Atas perhatiannya saya ucapkan terima kasih.

Wassalammu'alaikum wr. wb.

Yogyakarta, 16 Juli 2025

Pembimbing

Dr. Shofwatul Uyun, S.T., M.Kom.
NIP. 198205112006042002

LEMBAR PEDOMAN PENGGUNAAN TUGAS AKHIR

Tugas Akhir ini tidak dipublikasikan, tetapi tersedia di perpustakaan dalam lingkungan Universitas Islam Negeri (UIN) Sunan Kalijaga Yogyakarta, diperkenankan dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun, dan harus menyebutkan sumbernya sesuai dengan kebiasaan ilmiah. Dokumen Tugas Akhir ini merupakan hak milik UIN Sunan Kalijaga Yogyakarta.



PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI *ENSEMBLE*

IBNU RAJU HUMAM

21106050047

INTISARI

Klasifikasi data merupakan tugas mendasar dalam *machine learning* untuk mengenali pola. Namun, dengan beragamnya tipe data seperti numerik, kategorikal, dan campuran menjadi tantangan dalam pemilihan model yang optimal. Algoritma berbasis pohon seperti Decision Tree, termasuk teknik *ensemble* seperti Random Forest dan *boosting* seperti AdaBoost, Gradient Boosting, LightGBM, dan XGBoost, sering digunakan. Penelitian ini bertujuan untuk mengevaluasi pengaruh tipe data terhadap performa berbagai algoritma klasifikasi tersebut. Selain itu, penelitian ini juga mengevaluasi efektivitas seleksi fitur menggunakan LASSO dan optimalisasi *hyperparameter tuning*. Metodologi penelitian yang dilakukan yaitu dengan melakukan tiga skenario: (1) model dasar menggunakan seluruh fitur, (2) model dengan seleksi fitur LASSO, dan (3) model dengan LASSO yang dioptimalkan melalui *hyperparameter tuning*. Hasil penelitian menunjukkan bahwa algoritma *ensemble boosting* (Gradient Boosting, LightGBM, XGBoost) secara konsisten menunjukkan performa terbaik pada dataset numerik dan campuran. Di sisi lain, efektivitas optimalisasi melalui LASSO dan *tuning* menunjukkan hasil yang bervariasi, meskipun berpotensi meningkatkan F1-Score dan efisiensi komputasi, sering kali ditemukan adanya *trade-off* di antara keduanya. Evaluasi pada data murni kategorikal menghadapi keterbatasan akibat sulitnya menemukan dataset.

Kata Kunci: klasifikasi, *ensemble learning*, *boosting*, *bagging*, LASSO, *hyperparameter tuning*.

PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI *ENSEMBLE*

IBNU RAJU HUMAM

21106050047

ABSTRAK

Data classification is a fundamental task in machine learning for recognizing patterns. However, the diversity of data types, such as numerical, categorical, and mixed, poses a challenge in selecting the optimal model. Tree-based algorithms, such as Decision Trees, are frequently used, including ensemble techniques like Random Forest and boosting methods like AdaBoost, Gradient Boosting, LightGBM, and XGBoost. This study aims to evaluate the impact of data type on the performance of these classification algorithms. Additionally, the study assesses the effectiveness of feature selection using LASSO and hyperparameter tuning optimization. The research methodology involves comparing models across three scenarios: (1) a baseline model using all features, (2) a model with LASSO feature selection, and (3) a model with LASSO optimized through hyperparameter tuning. The results show that ensemble boosting algorithms (Gradient Boosting, LightGBM, XGBoost) consistently perform best on numerical and mixed datasets. On the other hand, the effectiveness of optimization through LASSO and tuning showed varying results. However, it has the potential to improve both the F1-Score and computational efficiency, as there is often a trade-off between the two. Evaluation of purely categorical data faces limitations due to the difficulty in finding suitable datasets.

Keywords: classification, ensemble learning, boosting, bagging, LASSO, hyperparameter tuning.

HALAMAN MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

اللَّهُمَّ صَلِّ عَلَى سَيِّدِنَا مُحَمَّدٍ وَعَلَى آلِ سَيِّدِنَا مُحَمَّدٍ

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٥﴾ إِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٦﴾

(Q.S. Al-Insyirah: 5-6)

"Keberhasilan bukanlah milik orang yang pintar. Keberhasilan adalah kepunyaan mereka yang senantiasa berusaha."

(Bacharuddin Jusuf Habibie)

"Kekuranganmu akan dianggap sempurna oleh hati yang ditakdirkan untukmu."

(Al Habib Umar bin Hafidz)

"Just keep swimming."

(Dory, Finding Nemo)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan terima kasih kepada Allah yang Maha Pengasih dan Maha Penyayang, saya persembahkan Tugas Akhir ini sebagai tanda penghormatan dan rasa cinta yang mendalam. Semoga karya ini dapat bermanfaat dan menjadi amal yang baik. Tugas Akhir ini saya persembahkan dengan penuh cinta dan hormat kepada:

1. Ayah dan Ibu tercinta, Ayah Anwar dan Ibu Roikanah, yang telah memberikan kasih sayang, doa, dan dukungan yang tiada henti. Engkau adalah sumber inspirasi dalam setiap langkah hidupku dan selalu menjadi sumber semangat tanpa henti.
2. Kedua kakakku, Sarrah Nur Layla dan Isnaini Nur Jannah, yang selalu menjadi sumber semangat dan mengisi hari-hari yang berwarna.
3. Semua sahabat dan teman seperjuangan yang telah memberikan dukungan dan semangat tanpa henti.

Terima kasih yang tak terhingga atas segala dukungan yang telah kalian berikan. Semoga Allah senantiasa melimpahkan rahmat dan berkah-Nya, serta memberikan balasan terbaik untuk semua kebaikan yang telah diberikan.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Alhamdulillah, segala puja dan puji kepada Allah Yang Maha Pengasih lagi Maha Penyayang, yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Selawat serta salam teruntuk Baginda Nabi Muhammad saw. semoga umatnya mendapatkan *syafa'at* beliau di *yaumul akhir* kelak. *Amin Ya Rabbal 'Alamin*.

Adapun tujuan utama dalam pembuatan laporan ini adalah sebagai Tugas Akhir, dan judul dari penelitian ini adalah **“PENGARUH KARAKTERISTIK DATA DAN OPTIMALISASI MODEL TERHADAP KINERJA ALGORITMA KLASIFIKASI *ENSEMBLE*”**.

Dengan penuh kerendahan hati, penulis ingin menyampaikan rasa terima kasih dan penghargaan kepada pihak-pihak berikut:

1. Kedua orang tua, Ayah dan Ibu, sumber doa, inspirasi, dan kekuatan tanpa batas. Semoga Allah Swt. senantiasa melimpahkan kesehatan, umur yang panjang dan berkah, serta kebahagiaan dunia dan akhirat.
2. Ayuk Sarrah, Ayuk Isni, Bang Ilham dan saudara-saudara lainnya yang selalu memberi dukungan untuk segera menyelesaikan tugas akhir ini.
3. Untuk diri sendiri, Ibnu Raju Humam. Terima kasih telah berjuang dan bertahan melewati segala tantangan, keraguan, dan lelah. Semua ini adalah bukti bahwa kamu mampu.
4. Untukmu, pasangan hidupku di masa depan. Semoga karya ini menjadi langkah awal dari cerita kebaikan yang akan kita bangun bersama.
5. Ibu Dr. Shofwatul Uyun, S.T., M.Kom., selaku Dosen Pembimbing. Terima kasih atas kesabaran, waktu, serta arahan yang mencerahkan dan penuh perhatian selama proses bimbingan.
6. Seluruh Bapak dan Ibu Dosen serta staf akademik Program Studi Informatika yang telah memberikan bekal ilmu pengetahuan yang sangat bermanfaat.
7. Untuk teman-teman pondok. Pertanyaan "Kapan selesai? Kapan sidang?" dari kalian adalah bentuk kepedulian yang tak ternilai.

8. Untuk teman-teman KKN: Ihsan, Digdo, Nindi, Rida, Maya, Fida, Nisaaur, Sekar, dan Ade. Terima kasih telah membuat 45 hari terasa mengenang sekali walaupun banyak sekali cobaan dan ujian yang mendatang.
9. Rekan-rekan seperjuangan Informatika angkatan 2021. Terkhusus kawan-kawan di tongkrongan SJ yang selalu memberi dukungan emosional, pendengar setia segala keluhan, dan sumber tawa di tengah penatnya penyusunan Tugas Akhir.
10. Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam pembuatan laporan Tugas Akhir ini masih jauh dari kata sempurna, mohon maaf apabila terdapat kesalahan dalam penulisan laporan ini. Penulis juga sangat menghargai kritik serta saran dari para pembaca demi bahan pertimbangan perbaikan laporan selanjutnya.

Yogyakarta, 16 Juli 2025

Penulis

Ibnu Raju Humam

NIM. 21106050047

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN TUGAS AKHIR	iii
SURAT PERSETUJUAN TUGAS AKHIR.....	iv
LEMBAR PEDOMAN PENGGUNAAN TUGAS AKHIR.....	v
INTISARI.....	vi
ABSTRAK	vii
HALAMAN MOTTO	viii
HALAMAN PERSEMBAHAN	ix
KATA PENGANTAR.....	x
DAFTAR ISI	xii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvi
DAFTAR PERSAMAAN	xvii
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	3
1.3 TUJUAN PENELITIAN	4
1.4 BATASAN PENELITIAN.....	4
1.5 MANFAAT PENELITIAN.....	5
1.6 SISTEMATIKA PENULISAN.....	5
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	7
2.1 TINJAUAN PUSTAKA	7
2.2 LANDASAN TEORI	10

2.2.1 MACHINE LEARNING	10
2.2.2 TEKNIK ENSEMBLE	11
2.2.3 CLASSIFIER.....	13
2.2.4 FEATURE SELECTION (LASSO).....	18
2.2.5 HYPERPARAMETER TUNING	19
2.2.6 METRIK EVALUASI.....	20
BAB III METODE PENELITIAN.....	23
3.1 LOKASI DAN WAKTU PENELITIAN	23
3.2 ALAT DAN BAHAN	23
3.3 DATA.....	23
3.4 DESAIN EKSPERIMENTAL DAN METODOLOGI EVALUASI	27
3.4.1 DESAIN TIGA SKENARIO	29
3.4.2 METRIK EVALUASI.....	30
3.4.3 VALIDASI MODEL.....	30
BAB IV HASIL DAN PEMBAHASAN.....	32
4.1 ANALISIS DATA.....	32
4.1.1 INFORMASI DATA	32
4.1.2 PREPROCESSING DATA	34
4.2 ANALISIS KOMPARATIF BERDASARKAN KARAKTERISTIK DATASET	39
4.2.1 KINERJA PADA DATASET NUMERIK.....	39
4.2.2 KINERJA PADA DATASET KATEGORIKAL	43
4.2.3 KINERJA PADA DATASET CAMPURAN.....	45
4.3 EVALUASI DAMPAK TEKNIK OPTIMALISASI.....	48
BAB V PENUTUP.....	55

5.1 KESIMPULAN	55
5.2 REKOMENDASI.....	56
5.3 SARAN	57
DAFTAR PUSTAKA	58
LAMPIRAN.....	65



DAFTAR TABEL

Tabel 2.1 Ringkasan Tinjauan Pustaka Penelitian Terkait	8
Tabel 4.1 Informasi Dataset Breast Cancer, Mushroom, Extrovert vs Introvert Behavior, dan Heart Failure	33
Tabel 4.2 Informasi Dataset Diabetes, Raisin Binary, Rice Type, dan Star Categorization	34
Tabel 4.3 Persentase Data Tersisa	36
Tabel 4.4 Parameter yang Dilakukan Penyetelan pada Setiap Algoritma	38
Tabel 4.5 F1-Score Dataset Numerik Skenario 1, Skenario 2, dan Skenario 3.....	40
Tabel 4.6 F1-Score Dataset Kategorikal Skenario 1, Skenario 2, dan Skenario 3	43
Tabel 4.7 F1-Score Dataset Campuran Skenario 1, Skenario 2, dan Skenario 3 ..	45
Tabel 4.8 Tingkat Reduksi Fitur oleh LASSO	48
Tabel 4.9 Persentase Peningkatan Nilai F1-Score dan Efisiensi Waktu Eksekusi terhadap setiap Skenario	50



DAFTAR GAMBAR

Gambar 2.1 Bagging	12
Gambar 2.2 Boosting	12
Gambar 2.3 Decision Tree.....	14
Gambar 2.4 Random Forest	15
Gambar 2.5 AdaBoost	16
Gambar 2.6 Gradient Boosting	16
Gambar 2.7 LightGBM	17
Gambar 2.8 XGBoost.....	18
Gambar 3.1 Proses Tahapan Penelitian	28
Gambar 4.1 Interquartile Range (IQR)	35
Gambar 4.2 Sebaran Kelas Target pada Tiap Dataset	37
Gambar 4.3 Visualisasi Salah Satu Pohon pada Model AdaBoost untuk Dataset Breast Cancer (Skenario 3)	42
Gambar 4.4 Visualisasi Salah Satu Pohon pada Model AdaBoost untuk Dataset Diabetes (Skenario 3).....	42
Gambar 4.5 Visualisasi Salah Satu Pohon pada Model LightGBM untuk Dataset Raisin Binary (Skenario 3).....	43
Gambar 4.6 Visualisasi Salah Satu Pohon pada Model XGBoost untuk Dataset Rice Type (Skenario 3).....	43
Gambar 4.7 Visualisasi Salah Satu Pohon pada Model LightGBM untuk Dataset Mushroom (Skenario 3)	44
Gambar 4.8 Visualisasi Salah Satu Pohon pada Model XGBoost untuk Dataset Personality (Skenario 3).....	47
Gambar 4.9 Visualisasi Salah Satu Pohon pada Model Gradient Boosting untuk Dataset Heart Failure (Skenario 3).....	47
Gambar 4.10 Visualisasi Salah Satu Pohon pada Model AdaBoost untuk Dataset Star (Skenario 3)	47

DAFTAR PERSAMAAN

Persamaan 2.1 <i>Accuracy</i>	20
Persamaan 2.2 <i>Precision</i>	20
Persamaan 2.3 <i>Recall</i>	20
Persamaan 2.4 <i>F1-Score</i>	20
Persamaan 2.5 <i>Execution Time</i>	20
Persamaan 4.1 <i>Peningkatan F1-Score</i>	49
Persamaan 4.2 <i>Speedup</i>	49



BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Dalam era digital yang semakin berkembang, data merupakan sebuah aset yang memiliki nilai ekonomi yang tinggi [1], [2]. Setiap hari, jumlah data yang dihasilkan terus meningkat, seperti data transaksi keuangan, data media sosial, sensor IoT, data kesehatan, dan sumber data lainnya. Data memiliki potensi besar untuk diolah dan dianalisis guna menghasilkan informasi yang bermanfaat dalam pengambilan keputusan [3]. Salah satu cara yang sering digunakan untuk menganalisis data adalah *machine learning*, khususnya dalam bidang klasifikasi. Klasifikasi dalam *machine learning* merupakan proses memprediksi kelas atau label dari suatu data berdasarkan fitur-fiturnya, dan telah diaplikasikan dalam berbagai bidang seperti kesehatan [4], [5], keuangan [6], pemasaran [7], dan lain-lain [8].

Dataset yang digunakan dalam praktik nyata sering kali memiliki karakteristik yang beragam, seperti ukuran yang besar, dimensi yang tinggi, ketidakseimbangan kelas (*imbalanced class*), atau campuran antara fitur numerik dan kategorikal, yang tentunya terdapat perlakuan yang berbeda. Selain itu, dataset tersebut mungkin mengandung fitur yang tidak relevan atau *redundan*, yang dapat menurunkan performa model klasifikasi. Fitur-fitur yang tidak relevan tidak hanya meningkatkan kompleksitas komputasi, tetapi juga dapat menyebabkan *overfitting*, mengurangi akurasi, penggunaan memori yang lebih besar, serta memperlambat proses analisis data [9], [10], bahkan model bisa menjadi terlalu spesifik terhadap *data training* dan gagal melakukan generalisasi dengan baik pada data baru. Untuk mengatasi masalah ini, *feature selection* seringkali menjadi langkah penting dalam *preprocessing data*.

Feature selection dalam *machine learning* adalah proses memilih subset fitur yang paling relevan dan informatif dari dataset asli, sehingga fitur-fitur yang tidak relevan dapat dihilangkan [11]. Fitur-fitur terpilih kemudian digunakan pada

tahapan klasifikasi. Dengan mengurangi jumlah fitur dapat membantu meningkatkan akurasi model, mengurangi waktu komputasi, dan memudahkan interpretasi model. *Feature selection* dibagi menjadi tiga, yakni *filter method*, *wrapped method*, dan *embedded method*. Setiap metode memiliki kelebihan dan kekurangan masing-masing, sehingga tidak ada yang optimal untuk setiap kasus.

Selain tantangan terkait fitur, pemilihan algoritma klasifikasi yang tepat juga menjadi kunci keberhasilan dalam menghasilkan model yang akurat dan *robust*, apalagi dalam konteks keilmuan yang bertujuan mencapai model dengan akurasi terbaik. Algoritma klasifikasi yang umum digunakan antara lain Decision Tree yang berbasis pohon. Namun, algoritma ini memiliki kelemahan yakni mudah mengalami *overfitting*, tidak stabil, sensitif terhadap perubahan data, dan akurasi tidak sebaik teknik lanjutannya yakni *bagging* dan *boosting* [12].

Untuk mengatasi kelemahan tersebut, teknik *ensemble learning* dikembangkan. *Ensemble learning* merupakan teknik menggabungkan beberapa model dasar atau *base learners* untuk meningkatkan akurasi dan ketahanan model. Teknik *ensemble* memiliki tiga bagian, yaitu *stacking*, *bagging* dan *boosting* (pada penelitian ini hanya akan membahas *bagging* dan *boosting*). *Bagging*, seperti yang diimplementasikan dalam algoritma Random Forest, yakni dengan cara melatih beberapa model secara independen dan menggabungkan hasilnya melalui *voting* atau *averaging*. Sementara itu, *boosting*, yang diimplementasikan dalam algoritma seperti AdaBoost, Gradient Boosting, dan XGBoost, bekerja dengan cara melatih model secara berurutan, di mana setiap model berusaha memperbaiki kesalahan model sebelumnya.

Meskipun teknik *ensemble* seperti *bagging* dan *boosting* telah banyak diaplikasikan, masih diperlukan penelitian untuk mengetahui bagaimana performa teknik-teknik ini dibandingkan dengan algoritma klasifikasi tradisional pada berbagai jenis dataset. Misalnya, bagaimana performa Random Forest atau XGBoost dibandingkan dengan Decision Tree pada dataset numerik? Atau bagaimana performanya pada dataset dengan fitur kategorikal? Selain itu, bagaimana pengaruh *feature selection* dan penyetelan *hyperparameter* terhadap

performa model-model tersebut? Pertanyaan-pertanyaan ini menjadi penting karena pemilihan algoritma dan fitur yang tepat dapat memengaruhi efektivitas dan efisiensi model klasifikasi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan performa algoritma klasifikasi Decision Tree dan teknik *ensemble* (bagging dan boosting) pada berbagai jenis dataset, serta mengevaluasi pengaruh *feature selection* dan *hyperparameter tuning* terhadap performa model. Dataset yang digunakan bervariasi, mulai dari dataset numerik, kategorikal, hingga dataset gabungan keduanya. Hasil penelitian ini diharapkan dapat memberikan pemahaman tentang algoritma dan teknik mana yang paling efektif untuk jenis data tertentu, serta sejauh mana *feature selection* dan *hyperparameter tuning* dapat meningkatkan performa model klasifikasi. Dengan demikian, penelitian ini diharapkan dapat menjadi acuan bagi peneliti dan praktisi dalam memilih metode klasifikasi dan teknik *preprocessing* dan optimalisasi yang sesuai.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah dijelaskan, beberapa permasalahan yang dapat dirumuskan dalam penelitian ini adalah:

1. Bagaimana performa algoritma Decision Tree dan teknik *ensemble* (bagging dan boosting) pada berbagai jenis dataset?
2. Apakah terdapat perbedaan signifikan dalam performa model ketika menggunakan dataset dengan karakteristik yang berbeda, seperti dataset numerik, kategorikal, atau gabungan keduanya?
3. Sejauh mana efektivitas seleksi fitur menggunakan *feature selection* (LASSO) dan *hyperparameter tuning* dalam meningkatkan nilai F1-Score dan efisiensi komputasi dari setiap algoritma yang diuji?

1.3 TUJUAN PENELITIAN

1. Membandingkan performa algoritma klasifikasi Decision Tree dengan teknik *ensemble* (*bagging* dan *boosting*) pada berbagai jenis dataset.
2. Menganalisis perbedaan performa model pada dataset dengan karakteristik yang berbeda, seperti dataset numerik, kategorikal, dan gabungan keduanya.
3. Memberikan rekomendasi tentang algoritma atau teknik yang paling efektif untuk setiap jenis dataset berdasarkan hasil evaluasi, teknik optimalisasi dan waktu komputasi yang dibutuhkan.

1.4 BATASAN PENELITIAN

Agar penelitian ini lebih terfokus, berikut adalah batasan yang diterapkan:

1. Jenis Algoritma: Penelitian ini hanya membandingkan algoritma klasifikasi Decision Tree dan teknik *ensemble bagging* dan *boosting*, tidak dengan *stacking* (Random Forest untuk *bagging* dan AdaBoost, Gradient Boosting, LightGBM serta XGBoost untuk *boosting*).
2. Jenis Dataset: Dataset yang digunakan terbatas pada dataset tabular numerik, kategorikal, atau gabungan keduanya dengan terdiri dari dua kelas klasifikasi disebut klasifikasi biner. Dataset gambar atau teks tidak termasuk dalam penelitian ini.
3. *Feature Selection*: Teknik *feature selection* yang digunakan terbatas pada metode tertentu, yaitu LASSO.
4. Metrik Evaluasi Utama: Meskipun penelitian ini mengukur *accuracy*, *precision*, *recall*, F1-Score, dan waktu komputasi, analisis perbandingan utama akan difokuskan pada F1-Score dan waktu komputasi. Pertimbangan ini karena F1-Score memberikan ukuran performa yang lebih seimbang untuk berbagai kondisi distribusi kelas, sementara waktu komputasi mengukur efisiensi praktis dari setiap model (Penjelasan lebih lanjut mengenai justifikasi ini dijabarkan pada Sub Bab 2.2.6).

5. Alat dan Bahasa Pemrograman: Penelitian ini menggunakan Python dengan pustaka seperti scikit-learn, XGBoost, dan Pandas untuk implementasi algoritma dan analisis data.

1.5 MANFAAT PENELITIAN

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis. Secara teoritis, penelitian ini akan memberikan pemahaman tentang perbandingan performa algoritma klasifikasi Decision Tree dengan teknik *ensemble bagging* dan *boosting* pada berbagai jenis dataset, serta menjelaskan pengaruh penerapan *feature selection* dan *hyperparameter tuning* terhadap performa model. Hasil penelitian juga diharapkan dapat menambah literatur dan referensi terkait pemilihan algoritma dan teknik *preprocessing* dan optimalisasi model yang tepat untuk dataset dengan karakteristik tertentu. Secara praktis, penelitian ini dapat memberikan rekomendasi tentang algoritma atau teknik yang paling efektif untuk setiap jenis dataset, membantu peneliti dan praktisi dalam memilih metode yang sesuai, serta meningkatkan akurasi dan efisiensi sistem klasifikasi dalam berbagai bidang seperti kesehatan, keuangan, dan pemasaran.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan berisi penjelasan gambaran umum yang akan dilakukan dalam penelitian dan penyusunan laporan. Struktur yang efektif merupakan faktor krusial dalam menyampaikan argumen dan temuan penelitian secara jelas dan sistematis. Berikut ini merupakan susunan dalam sistematika penulisan dalam penelitian ini.

BAB I PENDAHULUAN

Dalam bab ini dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, serta sistematika penulisan laporan penelitian.

BAB II TINJAUAN PUSTAKAN DAN LANDASAN TEORI

Dalam bab ini dipaparkan ringkasan penelitian-penelitian terdahulu yang berkaitan dengan penelitian ini. Selain itu, dijelaskan teori-teori dan metode yang digunakan dalam mendukung penelitian terkait dengan *feature selection* serta *ensemble*.

BAB III METODE PENELITIAN

Dalam bab ini dijelaskan bagaimana peneliti menerapkan metode yang digunakan dalam penelitian sesuai dengan alur penelitian yang telah dirancang.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini dijelaskan tentang hasil penelitian terkait klasifikasi menggunakan teknik *ensemble* dengan *bagging* dan *boosting*.

BAB V PENUTUP

Pada bab ini disampaikan kesimpulan dari penelitian yang telah dilakukan serta saran dari peneliti untuk pengembangan lebih lanjut dalam penelitian berikutnya.

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan analisis dan pembahasan hasil penelitian yang telah diuraikan pada bab sebelumnya, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Performa Algoritma pada Berbagai Karakteristik Data: Teknik *ensemble*, baik *bagging* (Random Forest) maupun *boosting* (AdaBoost, Gradient Boosting, LightGBM, XGBoost), secara konsisten menunjukkan kinerja yang jauh lebih unggul dibandingkan dengan algoritma klasifikasi Decision Tree pada semua jenis dataset yang diuji (numerik, kategorikal, dan campuran). Keunggulan ini utamanya disebabkan oleh kemampuan metode *ensemble* dalam mengurangi *overfitting* dan bias, yang merupakan kelemahan dari model Decision Tree tunggal. Algoritma *boosting* yang lebih modern, seperti Gradient Boosting, LightGBM dan XGBoost, sering kali menunjukkan performa terbaik pada dataset numerik dan campuran, menunjukkan efektivitas arsitektur mereka dalam menangani data yang heterogen dan kompleks.
2. Perbedaan Kinerja antar Jenis Dataset: Kinerja relatif dari setiap algoritma sangat dipengaruhi oleh karakteristik intrinsik dataset. Algoritma *boosting*, khususnya XGBoost dan LightGBM, menunjukkan keunggulan pada dataset numerik dan campuran yang kompleks karena kemampuannya dalam memodelkan hubungan non-linear yang rumit. Untuk dataset kategorikal dengan separabilitas kelas yang tinggi, metrik akurasi menjadi kurang relevan dalam penelitian, dan efisiensi komputasi menjadi faktor pembeda utama, di mana LightGBM menunjukkan performa yang sangat baik. Hal ini menunjukkan bahwa tidak ada satu algoritma yang universal digunakan untuk setiap kasus.
3. Efektivitas Teknik Optimalisasi (LASSO dan *Hyperparameter Tuning*): Efektivitas *feature selection* menggunakan LASSO dan *hyperparameter tuning* sangat bergantung pada karakteristik dataset.

- Seleksi fitur LASSO terbukti sangat efektif dalam meningkatkan efisiensi komputasi (waktu eksekusi) dengan mengurangi dimensionalitas data secara signifikan, terutama pada dataset dengan banyak fitur atau hasil dari *one-hot encoding* seperti Breast Cancer dan Personality. Meskipun sering kali meningkatkan F1-Score dengan menghilangkan *noise*, LASSO yang bersifat linear berisiko menghilangkan fitur yang penting dalam interaksi non-linear, yang terkadang dapat menurunkan performa pada model tertentu.
- *Hyperparameter tuning* tidak selalu menjamin peningkatan performa. Pada beberapa kasus, seperti pada model LGBM di dataset Heart Failure, optimalisasi memberikan peningkatan F1-Score yang sangat tinggi hingga 12,1%. Namun, pada kasus lain, proses ini justru dapat menurunkan kinerja atau secara signifikan memperlambat model karena beban komputasi tambahan, yang menandakan adanya *trade-off* antara potensi peningkatan akurasi dan biaya komputasi.

5.2 REKOMENDASI

Berdasarkan hasil kesimpulan, berikut adalah beberapa rekomendasi praktis bagi para peneliti dan praktisi di bidang *machine learning*:

1. Pemilihan Algoritma Awal: Untuk tugas klasifikasi pada data tabular (terutama numerik dan campuran), disarankan untuk memulai dengan algoritma *boosting* modern seperti Gradient Boosting, LightGBM atau XGBoost. Algoritma-algoritma ini secara konsisten memberikan performa dasar yang kuat dan sering kali menjadi yang terbaik bahkan dengan parameter *default*.
2. Penerapan *Feature Selection*: Pada dataset dengan dimensionalitas tinggi, penggunaan seleksi fitur seperti LASSO sangat dianjurkan. Teknik ini terbukti efektif mengurangi waktu komputasi secara signifikan dan berpotensi meningkatkan akurasi dengan mereduksi fitur yang tidak relevan.

3. Strategi Optimalisasi Model: *Hyperparameter tuning* sebaiknya tidak diterapkan secara sembarangan. Karena hasilnya terkadang memiliki *trade-off* antara nilai akurasi dan waktu komputasi. Praktisi perlu mempertimbangkan apakah potensi peningkatan performa yang marginal sepadan dengan biaya komputasi yang signifikan, terutama untuk aplikasi skala besar atau *real-time*.

5.3 SARAN

Penelitian ini telah memberikan wawasan tambahan, namun masih terdapat ruang untuk pengembangan di masa depan. Berikut adalah beberapa saran untuk penelitian selanjutnya:

1. Memperluas Cakupan Algoritma: Penelitian ini membatasi pada teknik *bagging* dan *boosting*. Penelitian selanjutnya dapat menyertakan metode *ensemble* lain seperti *stacking*, yang tidak dibahas pada penelitian ini.
2. Menggunakan Teknik Seleksi Fitur Lain: Penelitian ini hanya menggunakan LASSO. Akan sangat bermanfaat untuk membandingkan efektivitas LASSO dengan metode seleksi fitur dari kategori lain, seperti metode *wrapper* atau metode filter lainnya.
3. Investigasi Penanganan Kelas Tidak Seimbang: Penelitian ini sengaja tidak menerapkan teknik penyeimbangan kelas seperti SMOTE. Penelitian mendatang dapat secara khusus menganalisis hubungan antara algoritma *ensemble* dengan berbagai teknik *resampling* untuk melihat apakah kinerja pada dataset yang tidak seimbang (seperti Pima Indians Diabetes dan Heart Failure) dapat ditingkatkan.
4. Menggunakan Dataset yang Lebih Beragam: Meskipun menggunakan delapan dataset, penelitian selanjutnya dapat mengambil manfaat dari pengujian pada kumpulan data yang lebih besar dan lebih bervariasi untuk memperkuat generalisasi hasil kesimpulan. Begitu juga, dataset kategorikal yang terbatas dalam penelitian ini.

DAFTAR PUSTAKA

- [1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science* (80-.), vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.
- [2] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017, doi: 10.1109/ACCESS.2017.2696365.
- [3] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/ACCESS.2017.2756872.
- [4] A. K. Triantafyllidis and A. Tsanas, "Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature," *J. Med. Internet Res.*, vol. 21, no. 4, p. e12286, Apr. 2019, doi: 10.2196/12286.
- [5] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure," *Am. Heart J.*, vol. 229, pp. 1–17, Nov. 2020, doi: 10.1016/j.ahj.2020.07.009.
- [6] D. Hoang and K. Wiegratz, "Machine Learning Methods in Finance: Recent Applications and Prospects," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4293977.
- [7] J. Salminen, V. Yoganathan, J. Corporan, B. J. Jansen, and S.-G. Jung, "Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type," *J. Bus. Res.*, vol. 101, pp. 203–217, Aug. 2019, doi: 10.1016/j.jbusres.2019.04.018.
- [8] D. Maheshwari, B. Garcia-Zapirain, and D. Sierra-Sosa, "Quantum Machine Learning Applications in the Biomedical Domain: A Systematic Review,"

- IEEE Access*, vol. 10, pp. 80463–80484, 2022, doi: 10.1109/ACCESS.2022.3195044.
- [9] J. Abdollahi and B. Nouri-Moghaddam, “A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation,” *Iran J. Comput. Sci.*, vol. 5, no. 3, pp. 229–246, 2022, doi: 10.1007/s42044-022-00104-x.
- [10] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O’Sullivan, “A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction,” *Front. Bioinforma.*, vol. 2, no. June, pp. 1–17, 2022, doi: 10.3389/fbinf.2022.927312.
- [11] V. Bolón-Canedo and A. Alonso-Betanzos, *Recent Advances in Ensembles for Feature Selection*, 1st ed., vol. 147. in Intelligent Systems Reference Library, vol. 147. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-90080-3.
- [12] A. Plaia, S. Buscemi, J. Fürnkranz, and E. L. Mencía, “Comparing Boosting and Bagging for Decision Trees of Rankings,” *J. Classif.*, vol. 39, no. 1, pp. 78–99, Mar. 2022, doi: 10.1007/s00357-021-09397-2.
- [13] P. Ghosh *et al.*, “Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques,” *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [14] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, “A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities,” *Inf. Fusion*, vol. 64, pp. 205–237, Dec. 2020, doi: 10.1016/j.inffus.2020.07.007.
- [15] N. Biswas *et al.*, “Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques,” vol.

2023, 2023, doi: 10.1155/2023/6864343.

- [16] H. Jafarzadeh, M. Mahdianpari, E. Gill, F. Mohammadimanesh, and S. Homayouni, "Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation," *Remote Sens.*, vol. 13, no. 21, p. 4405, Nov. 2021, doi: 10.3390/rs13214405.
- [17] J. Nayak, P. B. Dash, and B. Naik, "An Advance Boosting Approach for Multiclass Dry Bean Classification," *J. Eng. Sci. Technol. Rev.*, vol. 12, no. 2, pp. 107–115, 2023, doi: 10.25103/jestr.162.14.
- [18] R. Mirzaeian, R. Nopour, Z. Asghari Varzaneh, M. Shafiee, M. Shanbehzadeh, and H. Kazemi-Arpanahi, "Which are best for successful aging prediction? Bagging, boosting, or simple machine learning algorithms?," *Biomed. Eng. Online*, vol. 22, no. 1, p. 85, Aug. 2023, doi: 10.1186/s12938-023-01140-9.
- [19] A. Panesar, *Machine Learning and AI for Healthcare*. Berkeley, CA: Apress, 2021. doi: 10.1007/978-1-4842-6537-6.
- [20] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [21] C. F. Tsai and Y. T. Sung, "Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches," *Knowledge-Based Syst.*, vol. 203, p. 106097, 2020, doi: 10.1016/j.knosys.2020.106097.
- [22] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020, doi: 10.14569/IJACSA.2020.0110277.
- [23] X. Zhao and X. Nie, "Splitting Choice and Computational Complexity

- Analysis of Decision Trees,” *Entropy*, vol. 23, no. 10, p. 1241, 2021, doi: 10.3390/E23101241.
- [24] Priyanka and D. Kumar, “Decision tree classifier: a detailed survey,” *Int. J. Inf. Decis. Sci.*, vol. 12, no. 3, pp. 246–269, 2020, doi: 10.1504/IJIDS.2020.10029122.
- [25] O. Günlük, J. R. Kalagnanam, M. Li, M. Menickelly, and K. Scheinberg, “Optimal decision trees for categorical data via integer programming,” *J. Glob. Optim.*, vol. 81, no. 1, pp. 233–260, 2021, doi: 10.1007/S10898-021-01009-Y.
- [26] G. Zhang and A. Gionis, “Regularized impurity reduction: accurate decision trees with complexity guarantees,” *Data Min. Knowl. Discov.*, vol. 37, no. 1, pp. 434–475, 2022, doi: 10.1007/s10618-022-00884-7.
- [27] M. Ibrahim, “Evolution of Random Forest from Decision Tree and Bagging: A Bias-Variance Perspective,” *Dhaka Univ. J. Appl. Sci. Eng.*, vol. 7, no. 1, pp. 66–71, Feb. 2023, doi: 10.3329/dujase.v7i1.62888.
- [28] B. D. M. and B. Narasimha Murthy, “Random Forest and Concept of Decision Tree Model,” in *Design of Experiments and Advanced Statistical Techniques in Clinical Research*, Singapore: Springer Singapore, 2020, pp. 133–156. doi: 10.1007/978-981-15-8210-3_3.
- [29] “Decision Trees, Random Forests and Boosting,” in *Introduction to Environmental Data Science*, Cambridge University Press, 2023, pp. 473–493. doi: 10.1017/9781107588493.015.
- [30] X. Li and K. Li, “Imbalanced data classification based on improved EIWAPSO-AdaBoost-C ensemble algorithm,” *Appl. Intell.*, vol. 52, no. 6, pp. 6477–6502, Apr. 2022, doi: 10.1007/s10489-021-02708-5.
- [31] W. Guo, Z. Luo, H. Chen, F. Hang, J. Zhang, and H. Al Bayatti, “AdaBoost Algorithm in Trustworthy Network for Anomaly Intrusion Detection,” *Appl. Math. Nonlinear Sci.*, vol. 8, no. 1, pp. 1819–1830, Jan. 2023, doi:

10.2478/amns.2022.2.0171.

- [32] Y. Ding, H. Zhu, R. Chen, and R. Li, “An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification,” *Appl. Sci.*, vol. 12, no. 12, p. 5872, Jun. 2022, doi: 10.3390/app12125872.
- [33] Y. Wang and L. Feng, “Improved Adaboost Algorithm for Classification Based on Noise Confidence Degree and Weighted Feature Selection,” *IEEE Access*, vol. 8, pp. 153011–153026, 2020, doi: 10.1109/ACCESS.2020.3017164.
- [34] X.-M. Wang, Y.-C. Ma, M.-H. Hsieh, and M.-H. Yung, “Quantum speedup in adaptive boosting of binary classification,” *Sci. China Physics, Mech. Astron.*, vol. 64, no. 2, p. 220311, Feb. 2021, doi: 10.1007/s11433-020-1638-5.
- [35] Z. Fang, S. Wang, and P. Perdikaris, “Ensemble learning for Physics Informed Neural Networks: a Gradient Boosting approach,” *arXiv.org*, vol. abs/2302.1, 2023, doi: 10.48550/arXiv.2302.13143.
- [36] M. Dong, L. Yao, X. Wang, B. Benatallah, S. Zhang, and Q. Z. Sheng, “Gradient Boosted Neural Decision Forest,” *IEEE Trans. Serv. Comput.*, vol. 16, pp. 330–342, 2023, doi: 10.1109/TSC.2021.3133673.
- [37] T. Thongthammachart, S. Araki, H. Shimadera, T. Matsuo, and A. Kondo, “Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO₂ and PM_{2.5} levels in Kansai region, Japan,” *Environ. Model. Softw.*, vol. 155, p. 105447, 2022, doi: 10.1016/j.envsoft.2022.105447.
- [38] T. Hanne, “Using the Light Gradient Boosting Machine for Prediction in QSAR Models,” *Algorithms Intell. Syst.*, pp. 99–111, 2023, doi: 10.1007/978-981-99-1435-7_10.
- [39] F. I. Kurniadi and P. D. Larasati, “Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke,” *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 6,

- no. 1, pp. 67–72, 2022, doi: 10.47970/siskom-kb.v6i1.328.
- [40] A. B. Sallow and S. Almufti, “eXtreme Gradient Boosting Algorithm with Machine Learning: a Review,” *Acad. J. Nawroz Univ.*, vol. 12, no. 2, pp. 320–334, 2023, doi: 10.25007/ajnu.v12n2a1612.
- [41] B. Sun, B. Sun, T. Sun, T. Sun, and P. Jiao, “Spatio-Temporal Segmented Traffic Flow Prediction with ANPRS Data Based on Improved XGBoost,” *J. Adv. Transp.*, vol. 2021, pp. 1–24, 2021, doi: 10.1155/2021/5559562.
- [42] A. Gouveia and M. Correia, “Network Intrusion Detection with XGBoost,” Chapman and Hall/CRC, 2020, pp. 137–166. doi: 10.1201/9780429270567-6.
- [43] S. Afrin *et al.*, “Supervised machine learning based liver disease prediction approach with LASSO feature selection,” *Bull. Electr. Eng. Informatics*, vol. 10, no. 6, pp. 3369–3376, 2021, doi: 10.11591/EEI.V10I6.3242.
- [44] R. G. Mantovani, T. Horváth, R. Cerri, S. B. Junior, J. Vanschoren, and A. C. P. L. F. de Carvalho, “An empirical study on hyperparameter tuning of decision trees,” *arXiv Learn.*, 2018, [Online]. Available: <http://export.arxiv.org/pdf/1812.02207>
- [45] P. P. Ippolito, “Hyperparameter Tuning,” 2022, pp. 231–251. doi: 10.1007/978-3-030-88389-8_12.
- [46] F. Arden and C. Safitri, “Hyperparameter Tuning Algorithm Comparison with Machine Learning Algorithms,” pp. 183–188, 2022, doi: 10.1109/ICITISEE57756.2022.10057630.
- [47] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, “Fast hyperparameter tuning using Bayesian optimization with directional derivatives,” *Knowl. Based Syst.*, vol. 205, p. 106247, 2020, doi: 10.1016/J.KNOSYS.2020.106247.
- [48] F. N. Rohman, F. Farikhin, and B. Surarso, “Hyperparameter Tuning of Random Forest Algorithm for Diabetes Classification,” *Int. J. Curr. Sci. Res.*

Rev., vol. 08, no. 01, 2025, doi: 10.47191/ijcsrr/v8-i1-31.

