

**IMPLEMENTASI ALGORITMA LIGHTGBM UNTUK DETEKSI DINI  
PENYAKIT JANTUNG**



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**Vincent Gunawan**  
**00000043802**

**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNIK DAN INFORMATIKA**  
**UNIVERSITAS MULTIMEDIA NUSANTARA**  
**TANGERANG**  
**2024**

**IMPLEMENTASI ALGORITMA LIGHTGBM UNTUK DETEKSI DINI  
PENYAKIT JANTUNG**



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

**Vincent Gunawan**

**0000043802**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Vincent Gunawan  
NIM : 00000043802  
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Skripsi saya yang berjudul:  
**Implementasi Algoritma LightGBM untuk Deteksi Dini Penyakit Jantung**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 16 Mei 2024



(Vincent Gunawan)

UMNI  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PERSETUJUAN

Skripsi dengan judul

### IMPLEMENTASI ALGORITMA LIGHTGBM UNTUK DETEKSI DINI PENYAKIT JANTUNG

oleh

Nama : Vincent Gunawan  
NIM : 00000043802  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah disetujui untuk diajukan pada

Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 16 Mei 2024

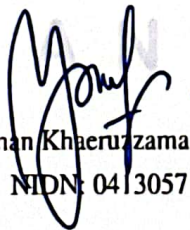
Dosen Pembimbing I



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

PJS Ketua Program Studi Informatika,



(Yaman/Khaeruzaman, M.Sc.)

NIDN: 0413057104

## HALAMAN PENGESAHAN

Skripsi dengan judul  
**IMPLEMENTASI ALGORITMA LIGHTGBM UNTUK DETEKSI DINI  
PENYAKIT JANTUNG**

oleh

Nama : Vincent Gunawan  
NIM : 00000043802  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Rabu, 29 Mei 2024

Pukul 13.00 s/s 15.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang



(David Agustriawan, S.Kom., M.Sc.,  
PhD.)

NIDN: 0525088601

Penguji



(Vincentius Kurniawan, S.Kom.,  
M.Eng.Sc.)

NIDN: 0308079501

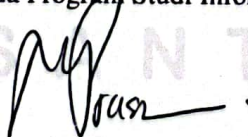
Pembimbing



(Dennis Gunawan, S.Kom., M.Sc.)

NIDN: 0320059001

Drs. Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

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

Yang bertanda tangan di bawah ini:

Nama : Vincent Gunawan  
NIM : 00000043802  
Program Studi : Informatika  
Jenjang : S1  
Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)\*\*.

Tangerang, 16 Mei 2024

Yang menyatakan

  
Vincent Gunawan

\*\* Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.



**Halaman Persembahan / Motto**

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)

UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma LightGBM untuk Deteksi Dini Penyakit Jantung 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. Ninok Leksono, 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 Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Dennis Gunawan, S.Kom., M.Sc., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Pihak - pihak dan teman lain yang telah membantu dalam pengerjaan proses skripsi secara moral maupun saran.

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

Tangerang, 16 Mei 2024



Vincent Gunawan



# IMPLEMENTASI ALGORITMA LIGHTGBM UNTUK DETEKSI DINI PENYAKIT JANTUNG

Vincent Gunawan

## ABSTRAK

Jantung manusia berperan sebagai organ vital yang memompa darah ke seluruh tubuh melalui empat ruang, yakni dua atrium dan dua ventrikel. Dengan begitu penyakit pada jantung akan sangat beresiko tinggi bagi penderitanya. Penyakit jantung merupakan penyakit penyebab utama kematian di dunia, yaitu sebesar 32%. Oleh karena itu, penting untuk dilakukan pendeteksian dini terhadap penderita penyakit jantung. Pendeteksian dini pada penyakit jantung terdapat beberapa metode medis seperti Elektrokardiogram dan *stress test* pada jantung. Dari hasil metode tersebut dapat dilakukan deteksi dini melalui pembelajaran mesin. Metode pembelajaran mesin yang digunakan adalah algoritma LightGBM yang ditingkatkan menggunakan *hyperparameter tuning*. Light Gradient Boosting Machine (LightGBM) merupakan kerangka kerja peningkatan gradien yang menggunakan algoritma pembelajaran berbasis pohon. Metode algoritma LightGBM berhasil diimplementasikan untuk melakukan deteksi dini penyakit jantung pada dataset *Heart Disease UCI Machine Learning* dengan performa accuracy sebesar 90.16%, precision sebesar 93.33%, recall sebesar 87.5%, F1-Score sebesar 90.32% dan AUC-ROC sebesar 94.07%. Hasil tersebut berdasarkan hasil terbaik dari 762.299 kombinasi *hyperparameter* yaitu learning rate sebesar 0.1, reg alpha sebesar 0.5, reg lambda sebesar 0.5, max depth sebesar 3, max estimator sebesar 100, dan number leaves sebesar 5.

**Kata kunci:** deteksi dini, *Gradient Boosting*, *Heart Disease UCI Machine Learning*, *LightGBM*, *Penyakit Jantung*

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

# Implementation of LightGBM Algorithm for Early Detection of Heart Disease

Vincent Gunawan

## ABSTRACT

*The human heart acts as a vital organ that pumps blood throughout the body through four chambers, namely two atria and two ventricles. Consequently, heart diseases pose a significant risk to affected individuals. Heart disease is the leading cause of death worldwide, accounting for 32% of all deaths. Therefore, early detection of heart disease is crucial. Several medical methods for early detection of heart disease include electrocardiograms and stress tests. The results from these methods can be further analyzed using machine learning techniques. One such technique is the LightGBM (Light Gradient Boosting Machine) algorithm, which is enhanced using hyperparameter tuning. LightGBM is a gradient boosting framework that uses tree-based learning algorithms. The LightGBM algorithm was successfully implemented to perform early detection of heart disease on the Heart Disease UCI Machine Learning dataset, achieving an accuracy of 90.16%, precision of 93.33%, recall of 87.5%, F1-Score of 90.32%, and AUC-ROC of 94.07%. These results were based on the best outcome from 762,299 hyperparameter combinations, with a learning rate of 0.1, reg alpha of 0.5, reg lambda of 0.5, max depth of 3, max estimators of 100, and number of leaves of 5.*

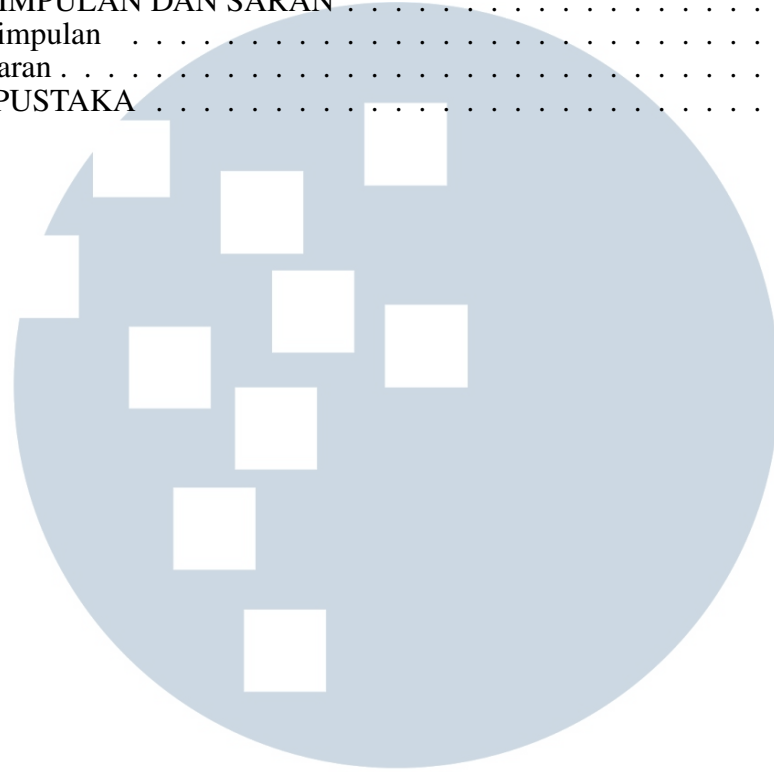
**Keywords:** *early detection, Gradient Boosting, Heart Disease, Heart Disease UCI Machine Learning, LightGBM*



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PERSETUJUAN . . . . .	iii
HALAMAN PENGESAHAN . . . . .	iv
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	v
HALAMAN PERSEMBAHAN/MOTO . . . . .	vi
KATA PENGANTAR . . . . .	vii
ABSTRAK . . . . .	viii
ABSTRACT . . . . .	ix
DAFTAR ISI . . . . .	x
DAFTAR GAMBAR . . . . .	xii
DAFTAR TABEL . . . . .	xiv
DAFTAR KODE . . . . .	xv
DAFTAR LAMPIRAN . . . . .	xvi
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	4
1.3 Batasan Permasalahan . . . . .	5
1.4 Tujuan Penelitian . . . . .	5
1.5 Manfaat Penelitian . . . . .	5
1.6 Sistematika Penulisan . . . . .	5
BAB 2 LANDASAN TEORI . . . . .	7
2.1 Penyakit Jantung . . . . .	7
2.2 <i>Machine Learning</i> . . . . .	7
2.3 <i>Supervised Learning</i> . . . . .	7
2.4 <i>Ensemble Learning</i> . . . . .	8
2.5 <i>Boosting</i> . . . . .	8
2.6 LightGBM . . . . .	9
2.7 <i>Confusion Matrix</i> . . . . .	14
BAB 3 METODOLOGI PENELITIAN . . . . .	16
3.1 Studi Literatur . . . . .	17
3.2 Pengumpulan Data . . . . .	17
3.3 Pengolahan Data . . . . .	19
3.3.1 <i>Single Imputation</i> . . . . .	20
3.3.2 Pembagian Data . . . . .	21
3.3.3 SMOTE ( <i>Synthetic Minority Oversampling Technique</i> ) . . . . .	21
3.3.4 Standarisasi Data . . . . .	22
3.4 Pembangunan Model . . . . .	22
3.5 Evaluasi Model . . . . .	24
3.6 Dokumentasi . . . . .	25
BAB 4 HASIL DAN DISKUSI . . . . .	26
4.1 Hasil Implementasi Pembelajaran Mesin . . . . .	26
4.1.1 Pengumpulan Data . . . . .	26
4.1.2 Pengolahan Data . . . . .	26
4.1.3 Pembangunan Model . . . . .	32
4.2 Skenario Perhitungan Algoritma LightGBM . . . . .	34
4.3 Uji Coba . . . . .	45
4.3.1 Skenario Uji Coba . . . . .	46

4.3.2	Hasil Uji Coba	47
4.4	Diskusi	58
BAB 5	SIMPULAN DAN SARAN	65
5.1	Simpulan	65
5.2	Saran	65
DAFTAR PUSTAKA		66



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

Gambar 2.1	Algoritma <i>Histogram-based</i> [1] . . . . .	10
Gambar 2.2	Algoritma <i>Gradient Based One Side Sampling</i> (GOSS)[1] . . . . .	11
Gambar 2.3	<i>Greedy Bundling</i> [1] . . . . .	12
Gambar 2.4	<i>Merge Exclusive Features</i> [1] . . . . .	13
Gambar 2.5	<i>Leaf-wise tree growth in LightGBM</i> [2] . . . . .	13
Gambar 2.6	<i>Level-wise tree growth in other boosting algorithms</i> [2] . . . . .	14
Gambar 2.7	<i>Confusion Matrix 2x2</i> [3] . . . . .	14
Gambar 3.1	<i>Alur Penelitian</i> . . . . .	16
Gambar 3.2	<i>Gambaran keseluruhan penelitian</i> . . . . .	17
Gambar 3.3	Dataset Heart Disease UCI Machine Learning Repository . . . . .	18
Gambar 3.4	<i>Corellation per feature</i> . . . . .	19
Gambar 3.5	<i>Flowchart Pengolahan Data</i> . . . . .	20
Gambar 3.6	<i>Flowchart Single Imputation</i> . . . . .	20
Gambar 3.7	<i>Flowchart Pembagian Data</i> . . . . .	21
Gambar 3.8	<i>Flowchart SMOTE</i> . . . . .	22
Gambar 3.9	<i>Flowchart Standarisasi Data</i> . . . . .	22
Gambar 3.10	<i>Flowchart Pembangunan Model</i> . . . . .	23
Gambar 3.11	<i>Flowchart Algoritma LightGBM</i> . . . . .	23
Gambar 3.12	<i>Flowchart Evaluasi Model</i> . . . . .	24
Gambar 4.1	<i>Import Dataset</i> . . . . .	26
Gambar 4.2	Mengubah <i>multi class</i> menjadi <i>binary</i> . . . . .	26
Gambar 4.3	Dataset dalam <i>multi class</i> . . . . .	27
Gambar 4.4	Dataset diubah ke <i>binary class</i> . . . . .	27
Gambar 4.5	Dataset yang memiliki value <i>null</i> . . . . .	28
Gambar 4.6	Jumlah data dengan value yang <i>null</i> . . . . .	28
Gambar 4.7	<i>Single Imputation</i> . . . . .	28
Gambar 4.8	Dataset setelah <i>single imputation</i> . . . . .	29
Gambar 4.9	Kode pembagian data <i>training</i> dan <i>test</i> dataset . . . . .	29
Gambar 4.10	Hasil pembagian data <i>train</i> . . . . .	29
Gambar 4.11	Hasil pembagian data <i>test</i> . . . . .	30
Gambar 4.12	Kode untuk melakukan SMOTE . . . . .	30
Gambar 4.13	Data sebelum dilakukan SMOTE . . . . .	31
Gambar 4.14	Data setelah dilakukan SMOTE . . . . .	31
Gambar 4.15	Kode Standarisasi data . . . . .	32
Gambar 4.16	Standarisasi Data Train . . . . .	32
Gambar 4.17	Standarisasi Data Test . . . . .	32
Gambar 4.18	Potongan kode <i>cross validation</i> . . . . .	33
Gambar 4.19	Potongan kode model LightGBM . . . . .	33
Gambar 4.20	Pembangunan Model LightGBM . . . . .	34
Gambar 4.21	Potongan kode pembagunan model . . . . .	34
Gambar 4.22	Graf <i>conflict</i> pada setiap <i>feature</i> . . . . .	35
Gambar 4.23	<i>Histogram</i> pada iterasi ke-1 . . . . .	38
Gambar 4.24	<i>Tree</i> pada iterasi ke-1 . . . . .	39
Gambar 4.25	<i>Tree</i> pada iterasi ke-1 . . . . .	40
Gambar 4.26	<i>Histogram</i> pada iterasi ke-2 . . . . .	42
Gambar 4.27	<i>Tree</i> pada iterasi ke-2 . . . . .	43
Gambar 4.28	<i>Tree</i> pada iterasi ke-2 . . . . .	44
Gambar 4.29	<i>Final model</i> . . . . .	44

Gambar 4.30	Hasil Skenario 1 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	48
Gambar 4.31	Hasil Skenario 1 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	48
Gambar 4.32	Hasil Skenario 2 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	49
Gambar 4.33	Hasil Skenario 2 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	50
Gambar 4.34	Hasil Skenario 3 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	51
Gambar 4.35	Hasil Skenario 3 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	51
Gambar 4.36	Hasil Skenario 4 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	52
Gambar 4.37	Hasil Skenario 4 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	53
Gambar 4.38	Hasil Skenario 5 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	54
Gambar 4.39	Hasil Skenario 5 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	54
Gambar 4.40	Hasil Skenario 6 ( <i>Accuracy</i> dan <i>Precision</i> ) . . . . .	55
Gambar 4.41	Hasil Skenario 6 ( <i>Recall</i> , <i>F1</i> dan <i>AUC-ROC</i> ) . . . . .	56
Gambar 4.42	<i>Accuracy learning curve</i> dengan 10 tahapan berdasarkan <i>hyperparameter</i> terbaik dari setiap kombinasi <i>hyperparameter</i> . . . . .	61
Gambar 4.43	<i>Loss learning curve</i> dengan 10 tahapan berdasarkan <i>hyperparameter</i> terbaik dari setiap kombinasi <i>hyperparameter</i> . . . . .	62
Gambar 4.44	<i>Accuracy learning curve</i> dengan 10 tahapan berdasarkan <i>hyperparameter</i> terbaik dari setiap kombinasi <i>hyperparameter</i> dengan <i>early stopping</i> . . . . .	63
Gambar 4.45	<i>Loss learning curve</i> dengan 10 tahapan berdasarkan <i>hyperparameter</i> terbaik dari setiap kombinasi <i>hyperparameter</i> dengan <i>early stopping</i> . . . . .	63





## DAFTAR TABEL

Tabel 1.1	Tabel perbandingan penelitian terkait dalam prediksi penyakit jantung . . . . .	3
Tabel 4.1	<i>Dataset awal</i> . . . . .	35
Tabel 4.2	<i>Bin Range</i> pada setiap feature data . . . . .	36
Tabel 4.3	Tahapan menggabungkan <i>feature</i> . . . . .	36
Tabel 4.4	Mencari <i>Gradien</i> dan <i>Hessian</i> dari dataset . . . . .	37
Tabel 4.5	Pembagian data ke dalam bin pada iterasi ke-1 . . . . .	37
Tabel 4.6	Mencari <i>split point</i> terbaik . . . . .	38
Tabel 4.7	Perhitungan GOSS . . . . .	40
Tabel 4.8	Mencari <i>Gradien</i> dan <i>Hessian</i> dari dataset iterasi ke-2 . . . . .	41
Tabel 4.9	Pembagian data ke dalam bin pada iterasi ke-2 . . . . .	42
Tabel 4.10	Mencari <i>split point</i> terbaik pada iterasi ke-2 . . . . .	43
Tabel 4.11	<i>Dataset Test</i> . . . . .	44
Tabel 4.12	Feature hasil EFB . . . . .	45
Tabel 4.13	<i>Hyperparameter</i> Terbaik berdasarkan score AUC-ROC tertinggi . . . . .	46
Tabel 4.14	<i>Hyperparameter</i> Default LightGBM . . . . .	46
Tabel 4.15	Tabel Skenario Uji Coba . . . . .	47
Tabel 4.16	Hasil terbaik dari setiap skenario berdasarkan <i>hyperparameter default</i> . . . . .	56
Tabel 4.17	Kombinasi <i>hyperparameter</i> terbaik dari semua kombinasi parameter . . . . .	57
Tabel 4.18	Confusion Matrix algoritma LightGBM terbaik berdasarkan semua kombinasi parameter . . . . .	57
Tabel 4.19	Hasil Performa algoritma LightGBM terbaik dalam semua kombinasi <i>hyperparameter</i> . . . . .	58
Tabel 4.20	Perbandingan harga <i>test</i> jantung yang dilakukan di Indonesia . . . . .	58



## DAFTAR KODE

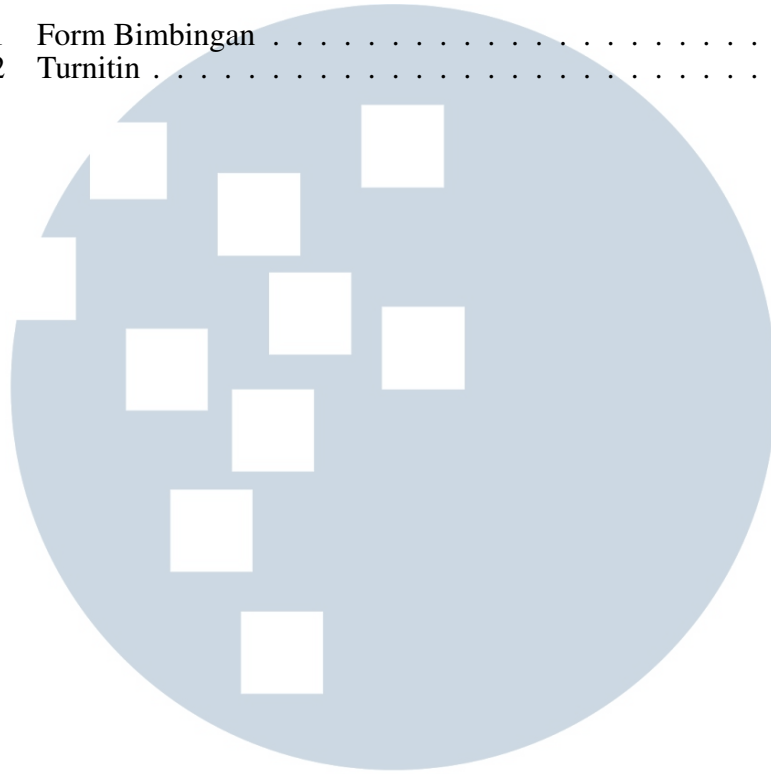


# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan . . . . .	70
Lampiran 2	Turnitin . . . . .	71



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA