SKRIPSI

OPTIMASI EXTREME GRADIENT BOOSTING MENGGUNAKAN SMOTE-ENN DAN HYPERPARAMETER TUNING OPTUNA UNTUK PREDIKSI PENYAKIT JANTUNG

MUTYA ANISSA MAULINDA

NPM 21081010154

DOSEN PEMBIMBING

Dr. Budi Nugroho, S.Kom., M.Kom.

Achmad Junaidi, S.Kom., M.Kom.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2026

LEMBAR PENGESAHAN

OPTIMASI EXTREME GRADIENT BOOSTING MENGGUNAKAN SMOTE-ENN DAN HYPERPARAMETER TUNING OPTUNA UNTUK PREDIKSI PENYAKIT JANTUNG

Oleh :
MUTYA ANISSA MAULINDA
NPM. 21081010154

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakulktas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 20 Agustus 2026.

Dr. Budi Nugroho, S.Kom., M.Kom.
NIP. 19800907 202121 1 005
..... (Pembimbing I)

Achmad Junaidi, S.Kom., M.Kom
NIP. 19781110 202521 1 048
..... (Pembimbing II)

Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 202121 2 005
..... (Ketua Penguji)

Eva Yulia Puspaningrum, S.Kom., M.Kom
NIP. 19890705 202121 2 002
..... (Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

OPTIMASI EXTREME GRADIENT BOOSTING MENGGUNAKAN SMOTE-ENN DAN HYPERPARAMETER TUNING OPTUNA UNTUK PREDIKSI PENYAKIT JANTUNG

Oleh :
MUTYA ANISSA MAULINDA
NPM. 21081010154

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Informatika
Fakultas Ilmu Komputer

Dr. Intan Yuniar Purbasari, S.Kom. MSc.
NIP. 19800602 202521 2 029

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama	: Mutya Anissa Maulinda
NPM	: 21081010154
Program	: Sarjana(S1)
Program Studi	: Informatika
Fakultas	: Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemulan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 13 November 2026

Yang Membuat pernyataan



Mutya Anissa Maulinda
NPM. 21081010154

ABSTRAK

Nama Mahasiswa / NPM : Mutya Anissa Maulinda / 21081010154

Judul Skripsi : Optimasi Extreme Gradient Boosting Menggunakan SMOTE-ENN dan Hyperparameter Tuning Optuna untuk Prediksi Penyakit Jantung

Dosen Pembimbing : 1. Dr. Budi Nugroho, S.Kom. M.Kom

2. Achmad Junaidi, S.Kom. M.Kom

ABSTRAK

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, sehingga pengembangan model prediksi yang akurat menjadi penting dalam bidang *machine learning*. Kinerja model dapat dipengaruhi oleh perbedaan distribusi kelas dan pemilihan *hyperparameter* yang belum optimal. Penelitian ini bertujuan meningkatkan kinerja XGBoost dalam memprediksi penyakit jantung melalui penerapan *Synthetic Minority Oversampling Technique-Edited Nearest Neighbors* (SMOTE-ENN) dan optimasi *hyperparameter* menggunakan Optuna berbasis *Tree-structured Parzen Estimator* (TPE). Dataset yang digunakan berasal dari UCI Machine Learning Repository. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, F1-score, dan *Area Under the Curve* (AUC) pada beberapa rasio pembagian data dan *k-fold cross-validation*. Hasil menunjukkan bahwa kombinasi SMOTE-ENN dan Optuna menghasilkan performa terbaik dengan rata-rata *accuracy* 97,65%, *precision* 97,52%, *recall* 97,34%, F1-*score* 97,43%, dan AUC 99,77%. Hasil tersebut menunjukkan bahwa penerapan SMOTE-ENN dan optimasi *hyperparameter* Optuna dapat meningkatkan kinerja XGBoost dalam prediksi penyakit jantung.

Kata kunci: Prediksi Penyakit Jantung, XGBoost, SMOTE-ENN, Optuna, Optimasi Hyperparameter.

ABSTRACT

Student Name / NPM : *Mutya Anissa Maulinda / 21081010154*

Thesis Title : *Optimizing Extreme Gradient Boosting Using SMOTE-ENN and Optuna-Based Hyperparameter Optimization for Heart Disease Prediction*

Advisor : *1.Dr. Budi Nugroho, S.Kom. M.Kom*

2. Achmad Junaidi, S.Kom.M.Kom

ABSTRACT

Heart disease is one of the leading causes of death worldwide, making the development of accurate predictive models important in the field of machine learning. Model performance can be affected by differences in class distribution and suboptimal hyperparameter selection. This study aims to improve the performance of XGBoost for heart disease prediction through the application of Synthetic Minority Oversampling Technique-Edited Nearest Neighbors (SMOTE-ENN) and hyperparameter optimization using Optuna based on the Tree-structured Parzen Estimator (TPE). The dataset used in this study was obtained from the UCI Machine Learning Repository. Model evaluation was performed using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) across several data-splitting ratios and k-fold cross-validation settings. The results showed that the combination of SMOTE-ENN and Optuna achieved the best performance, with average accuracy 97,65%, *precision* 97,52%, *recall* 97,34%, F1-*score* 97,43%, dan AUC 99,77%. *These results indicate that the application of SMOTE-ENN combined with Optuna hyperparameter optimization can improve the performance of XGBoost for heart disease prediction.*

Keywords: *Heart disease prediction, XGBoost, SMOTE-ENN, Optuna Hyperparameter optimization.*

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Optimasi Extreme Gradient Boosting Menggunakan SMOTE-ENN dan Hyperparameter Tuning Optuna untuk Prediksi Penyakit Jantung" dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika.

Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua dan keluarga yang senantiasa memberikan doa, kasih sayang, motivasi, serta dukungan moral maupun material kepada penulis.
2. Ibu Prof. Dr. Ir. Rossyda Priyadarshini, MP., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Intan Yuniar Purbasari, S.Kom. MSc, selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Made Hanindia Prami Swari, S.Kom, M.Cs, selaku Dosen Wali yang telah memberikan arahan, bimbingan, serta motivasi.
6. Bapak Dr. Budi Nugroho, S.Kom, M.Kom selaku Dosen Pembimbing pertama dan Bapak Achmad Junaidi, S.Kom. M.Kom, yang telah memberikan bimbingan, saran, dan masukan yang sangat bermanfaat sehingga skripsi ini dapat diselesaikan dengan baik.
7. Seluruh dosen Program Studi Informatika yang telah memberikan ilmu, pengalaman, dan bimbingan selama masa perkuliahan.
8. Teman-teman dan semua pihak yang telah memberikan bantuan, semangat, dan dukungan sehingga skripsi ini dapat terselesaikan.
9. Semua pihak yang telah membantu dan mendukung penulis selama proses penyusunan skripsi ini, yang tidak dapat disebutkan satu per satu.

Penulis memahami bahwa karya ilmiah ini masih memiliki kekurangan yang perlu diperbaiki. Masukan berupa kritik dan saran yang konstruktif diharapkan dapat menjadi pertimbangan dalam pengembangan penelitian berikutnya. Semoga hasil penelitian ini dapat memperluas pemahaman pembaca serta menjadi referensi yang berguna bagi kajian ilmiah di bidang terkait.

Surabaya, 30 Oktober 2026

Penulis

DAFTARISI

LEMBAR PENGESAHAN ..... **i**
LEMBAR PERSETUJUAN ..... **iii**
SURAT PERNYATAAN ORISINALITAS ..... **v**
ABSTRAK ..... **vii**
ABSTRACT ..... **ix**
KATA PENGANTAR ..... **xi**
DAFTAR ISI ..... **xiii**
DAFTAR GAMBAR ..... **xvi**
DAFTAR TABEL ..... **xviii**
BAB I PENDAHULUAN **1**
1.1. Latar Belakang 1
1.2. Rumusan Masalah 5
1.3. Tujuan Penelitian 5
1.4. Manfaat Penelitian 5
1.5. Batasan Masalah 6
BAB II KAJIAN PUSTAKA **7**
2.1. Penelitian Terdahulu 7
2.2. Landasan Teori 13
2.2.1. Penyakit Jantung 13
2.2.2. *DataMining* 14
2.2.3. *Machine Learning* 14
2.2.4. *Imbalance Data* 15
2.2.5. Binarisasi Target 15
2.2.6. *Missing Value* dan Imputasi *Mean* 16
2.2.7. Normalisasi *Min-Max* 17
2.2.8. *Synthetic Minority Over-sampling Technique–Edited Nearest Neighbors* (SMOTE-ENN) 19
2.2.9. *Extreme Gradient Boosting* (XGBoost) 22
2.2.10. Optuna 26
2.2.11. *Confusion Matrix* 27
BAB III METODOLOGI **30**
3.1. Metode Penelitian 30

3.2. Tahapan Penelitian 30
3.3. Studi Literatur 31
3.4. Identifikasi Masalah 31
3.5. Pengumpulan Data 32
3.6. *Preprocessing* 34
3.6.1 *Encoding* Target Dataset Statlog 35
3.6.2 Binarisasi Target 35
3.6.3 Nilai Hilang (*Missing Value*) 36
3.6.4 Normalisasi Dataset 39
3.6.5 Seleksi Fitur *Chi-Square* 41
3.7. Synthetic Minority Over-sampling Technique–Edited Nearest Neighbors (SMOTE-ENN) 43
3.8. Extreme Gradient Boosting (XGBoost) 45
3.9. Optuna 51
3.10. Skenario Pengujian 53
BAB IV HASIL DAN PEMBAHASAN 56
4.1 *Preprocessing Data* 56
4.1.1 Pengambilan Dataset 56
4.1.2 *Encoding* Dataset Statlog 57
4.1.3 Penggabungan Dataset 58
4.1.4 *Missing Value* 59
4.1.5 Binarisasi Target 60
4.1.6 Imputasi (*mean* dan *modus*) 61
4.1.7 Normalisasi (*Min-Max Scaler*) 64
4.1.8 Seleksi Fitur *Chi-Square* 65
4.2 Implementasi Skenario Pengujian 66
4.2.1 Skenario XGBoost *Baseline* 66
4.2.1.1 Pembagian Data Latih dan Data Uji 67
4.2.1.2 Pelatihan Model XGBoost 67
4.2.1.3 Prediksi Data Uji 68
4.2.2 Skenario XGBoost dan SMOTE-ENN 69
4.2.2.1 Penerapan SMOTE-ENN 69
4.2.2.2 Pembagian Data Latih dan Data Uji 72
4.2.2.3 Pelatihan Model XGBoost 73
4.2.2.4 Prediksi Data Uji 73

4.2.3 Skenario XGBoost dan Optuna 74
4.2.3.1 Pembagian Data Latih dan Data Uji 74
4.2.3.2 Pelatihan XGBoost dengan Optimasi Optuna 75
4.2.3.3 Prediksi Data Uji 77
4.2.4 Skenario XGBoost, SMOTE-ENN, dan Optuna 78
4.2.4.1 Penerapan SMOTE-ENN 78
4.2.4.2 Pembagian Data Latih dan Data Uji 79
4.2.4.3 Pelatihan Model XGBoost, SMOTE-ENN, dan Optuna 80
4.2.4.4 Prediksi Data Uji 81
4.3 Pengujian 82
4.3.5 Analisis Hasil Evaluasi Skenario 83
BAB V KESIMPULAN DAN SARAN 90
5.1 Kesimpulan 90
5.2 Saran 90
DAFTARPUSTAKA 91

DAFTARGAMBAR

Gambar 2. 1 Ilustrasi SMOTE 20
Gambar 2. 2 Ilustrasi SMOTE-ENN 21
Gambar 2. 3 Ilustrasi XGBoost 22
Gambar 2. 4 Diagram Alur Optuna 26
Gambar 2. 5 *Confussion Matrix* 28
Gambar 3. 1 Kerangka Penelitian 30
Gambar 3. 2 Flowchart Pembuatan Model 32
Gambar 3. 3 Flowchart *Missing Value* 36
Gambar 3. 4 Flowchart Normalisasi 39
Gambar 3. 5 Flowchart SMOTE-ENN 43
Gambar 3. 6 Flowchart XGBoost 45
Gambar 4. 1 Dataset Sebelum Nama Kolom Didefinisikan 57
Gambar 4. 2 Dataset Setelah Nama Kolom Didefinisikan 57
Gambar 4. 3 Distribusi Target Statlog *encoding* 58
Gambar 4. 4 Penggabungan Dataset 59
Gambar 4. 5 Dataset Yang Mengandung Nilai "?" 60
Gambar 4. 6 Dataset setelah konversi NaN 60
Gambar 4. 7 Dataset sebelum binarisasi 61
Gambar 4. 8 Dataset setelah binarisasi 61
Gambar 4. 9 Jumlah *Missing Value* 63
Gambar 4. 10 Dataset yang telah diimputasi *Mean* 63
Gambar 4. 11 Dataset setelah normalisasi *Min-Max Scaling* 65
Gambar 4. 12 Hasil Chi-Square 66
Gambar 4. 13 Distribusi Kelas SMOTE-ENN 70
Gambar 4. 14 Visualisasi Distribusi Kelas 71
Gambar 4. 15 Visualisasi Proses SMOTE-ENN 71
Gambar 4. 16 Kurva Konvergensi Optuna 77
Gambar 4. 17 Grafik Matrix Accuracy 85
Gambar 4. 18 Grafik Matrix Precision 86
Gambar 4. 19 Grafik Matrix Recall 87
Gambar 4. 20 Grafik Matrix F1-score 87

Gambar 4. 21 Grafik Matrix AUC.....88
Gambar 4. 22 Kurva ROC.....89

DAFTAR TABEL

Tabel 3. 1 Distribusi Dataset 33
Tabel 3. 2 Deskripsi Atribut Dataset 33
Tabel 3. 3 Konversi Label Target Statlog 35
Tabel 3. 4 Dataset Binarisasi 35
Tabel 3. 5 *Data Missing Values* 36
Tabel 3. 6 Data Setelah Konversi NaN 37
Tabel 3. 7 Data Imputasi *Mean* dan *Modus* 38
Tabel 3. 8 Variabel yang Dinormalisasi 39
Tabel 3. 9 Data Sebelum Normalisasi 40
Tabel 3. 10 Data Setelah Normalisasi 41
Tabel 3. 11 Parameter yang Diuji 52
Tabel 3. 12 Skenario Pengujian Split Data Dan Fold 55
Tabel 4. 1 Matrix Evaluasi 82
Tabel 4. 2 Ringkasan Hasil Evaluasi Performa Tiap Skenario 83