



SKRIPSI

OPTIMASI RANDOM FOREST MENGGUNAKAN KOMBINASI SMOTE-ENN DAN SELEKSI FITUR RFECV UNTUK PREDIKSI PENYAKIT JANTUNG

SABRINA PUTRI AULIA

NPM 21081010048

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., M.T

Achmad Junaidi, S.Kom., M.Kom

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025



SKRIPSI

OPTIMASI RANDOM FOREST MENGGUNAKAN KOMBINASI SMOTE-ENN DAN SELEKSI FITUR RFECV UNTUK PREDIKSI PENYAKIT JANTUNG

SABRINA PUTRI AULIA

NPM 21081010048

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., M.T

Achmad Junaidi, S.Kom., M.Kom

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025

LEMBAR PENGESAHAN

OPTIMASI RANDOM FOREST MENGGUNAKAN KOMBINASI SMOTE-ENN DAN SELEKSI FITUR RFECV UNTUK PREDIKSI PENYAKIT JANTUNG

Oleh:

SABRINA PUTRI AULIA
NPM. 21081010048

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 16 Mei 2025

Menyetujui

Dr. Basuki Rahmat, S.Si., M.T.
NIP. 19690723 202121 002


..... (Pembimbing I)

Achmad Junaidi, S.Kom., M.Kom.
NPT. 3 7811 04 0199 1


..... (Pembimbing II)

Budi Nugroho, S.Kom. M.Kom.
NIP. 19800907 202121 1 005


..... (Ketua Penguji)

Eva Yulia Puspaningrum, S.Kom., M.Kom.
NIP. 19890705 202121 2 002


..... (Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer


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

Halaman ini sengaja dikosongkan

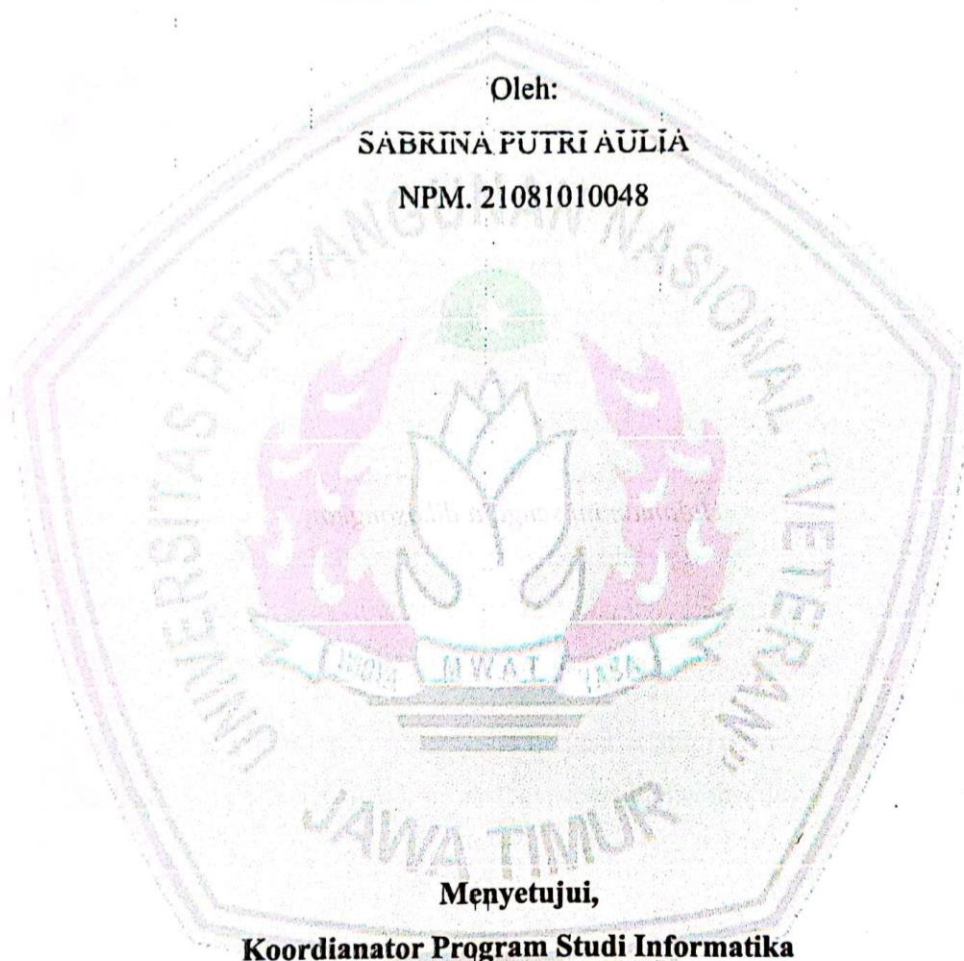
LEMBAR PERSETUJUAN

**OPTIMASI RANDOM FOREST MENGGUNAKAN KOMBINASI
SMOTE-ENN DAN SELEKSI FITUR RFECV UNTUK
PREDIKSI PENYAKIT JANTUNG**

Oleh:

SABRINA PUTRI AULIA

NPM. 21081010048



Menyetujui,

Koordinator Program Studi Informatika

Fakultas Ilmu Komputer

A handwritten signature in blue ink, appearing to read 'Fetty Tri Anggraeny', is written over a horizontal line.

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertandatangan dibawah ini:

Nama Mahasiswa : Sabrina Putri Aulia
NPM : 21081010048
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 plagiarisme. Apabila adikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerimasanksi 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, 16 Mei 2025
Yang Membuat Pernyataan,



SABRINA PUTRI AULIA
NPM. 21081010048

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Sabrina Putri Aulia / 21081010048
Judul Skripsi : Optimasi Random Forest Menggunakan
Kombinasi SMOTE-ENN dan Seleksi Fitur
RFECV untuk Prediksi Penyakit Jantung
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si., M.T
2. Achmad Junaidi, S.Kom., M.Kom

Penyakit jantung merupakan penyebab kematian utama secara global, dengan dampak signifikan terhadap beban ekonomi dan produktivitas masyarakat. Klasifikasi penyakit jantung diperlukan untuk mendukung deteksi dini dan penanganan lebih cepat, sehingga dapat menurunkan risiko komplikasi yang fatal. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi penyakit jantung melalui optimasi algoritma Random Forest menggunakan kombinasi metode Synthetic Minority Over-sampling Technique–Edited Nearest Neighbors (SMOTE-ENN) dan seleksi fitur Recursive Feature Elimination with Cross Validation (RFECV). Penelitian ini membandingkan empat skenario berbeda, yaitu tanpa SMOTE-ENN dan RFECV, hanya dengan SMOTE-ENN, hanya dengan RFECV, serta kombinasi SMOTE-ENN dan RFECV. Evaluasi dilakukan pada rasio pembagian data latih dan uji (70:30, 75:25, dan 80:20) dengan variasi parameter rasio sampling dan jumlah tetangga pada SMOTE-ENN serta jumlah lipatan pada RFECV. Hasil terbaik diperoleh pada kombinasi SMOTE-ENN dengan rasio sampling 1.0 dan 5 tetangga, serta RFECV dengan 3-Fold pada rasio data 80:20. Model menghasilkan akurasi 95.7%, precision 89.6%, recall 98.4%, dan F1-score 93.8%, yang menunjukkan bahwa kombinasi metode ini dapat meningkatkan kinerja Random Forest dalam klasifikasi penyakit jantung.

Kata kunci: Penyakit Jantung, Random Forest, SMOTE-ENN, RFECV, Prediksi, Pembelajaran Mesin.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Sabrina Putri Aulia / 21081010048
Thesis Title : Random Forest Optimization Using a
Combination of SMOTE-ENN and RFECV
Feature Selection
for Heart Disease Prediction
Advisor : 1. Dr. Basuki Rahmat, S.Si., M.T
2. Achmad Junaidi, S.Kom., M.Kom

Heart disease is the leading cause of death globally, with a significant impact on the economic burden and productivity of society. Heart disease classification is needed to support early detection and faster treatment, so as to reduce the risk of fatal complications. This study aims to improve the accuracy of heart disease classification through Random Forest algorithm optimization using a combination of the Synthetic Minority Over-sampling Technique-Edited Nearest Neighbors (SMOTE-ENN) method and Recursive Feature Elimination with Cross Validation (RFECV) feature selection. This study compares four different scenarios, namely without SMOTE-ENN and RFECV, only with SMOTE-ENN, only with RFECV, and a combination of SMOTE-ENN and RFECV. The evaluation was carried out using different training and testing data ratios (70:30, 75:25, and 80:20), with variations in the sampling ratio and number of neighbors for SMOTE-ENN, as well as the number of folds in RFECV. The best results were obtained with SMOTE-ENN using a sampling ratio of 1.0 and 5 neighbors, combined with 3-Fold RFECV on an 80:20 data split. The model achieved an accuracy of 95.7%, precision of 89.6%, recall of 98.4%, and F1-score of 93.8%, indicating that this combination improves the performance of Random Forest in heart disease classification.

Keywords: Heart Disease, Random Forest, SMOTE-ENN, RFECV, Prediction, Machine Learning

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Segala puji dan syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga skripsi penulis dengan judul **“Optimasi Random Forest Menggunakan Kombinasi SMOTE-ENN dan Seleksi Fitur RFECV untuk Prediksi Penyakit Jantung”** dapat diselesaikan dengan baik. Laporan ini disusun dan diajukan sebagai salah satu syarat untuk menyelesaikan program Strata Satu (S1) pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur.

Penulis menyadari sepenuhnya bahwa telah banyak mendapatkan dukungan, bimbingan dan dorongan dari berbagai pihak. Maka, pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT., IPU, selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom selaku Kepala Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Budi Nugroho, S.Kom, M.Kom selaku Dosen Wali yang telah banyak memberikan arahan dan masukan kepada penulis dalam masa perkuliahan.
5. Bapak Dr. Basuki Rahmat, S.Si., M.T dan Bapak Achmad Junaidi, S.Kom., M.Kom selaku dosen pembimbing yang telah memberikan arahan dan masukan selama proses penyusunan skripsi ini hingga selesai.
6. Teman-teman yang memberikan dukungan dan kerja samanya sehingga skripsi ini dapat terselesaikan dengan baik.
7. Rekan dan teman-teman penulis lainnya di luar perkuliahan yang memberikan dukungan untuk penulis untuk menyelesaikan program.
8. Seluruh keluarga yang senantiasa mendo’akan dan mendukung dalam setiap masa perkuliahan penulis.

9. Dan untuk semua pihak yang telah membantu hingga terselesaikan pembuatan Laporan Skripsi yang tentunya tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat berbagai kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun dari berbagai pihak guna menyempurnakan penulisan skripsi ini. Pada akhirnya, dengan segala keterbatasan yang dimiliki, penulis berharap laporan ini dapat memberikan manfaat bagi semua pihak secara umum, serta bagi penulis sendiri secara khusus.

Surabaya, 16 Mei 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	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	3
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu.....	7
2.2. Landasan Teori	11
2.2.1. Penyakit Jantung	11
2.2.2. Data Mining	12
2.2.3. Machine Learning	12
2.2.4. Imputasi Hot-Deck	13
2.2.5. Min-Max Scalling	13

2.2.6.	Balancing Data SMOTE-ENN	14
2.2.7.	Recursive Feature Elimination with Cross Validation (RFECV) ..	17
2.2.8.	Random Forest.....	17
2.2.9.	Confusion Matrix.....	20
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....		23
3.1.	Metode Penelitian	23
3.2.	Desain Sistem	23
3.2.1.	Pengumpulan Data.....	25
3.2.2.	Pra Proses Data.....	27
3.2.3.	Balancing Data dengan SMOTE-ENN.....	30
3.2.4.	Seleksi Fitur dengan RFECV	34
3.2.5.	Implementasi Model Random Forest	36
3.2.6.	Skema Pengujian	41
BAB IV PENGUJIAN DAN ANALISA.....		45
4.1.	Pra Proses Data.....	45
4.1.1.	Pengambilan dan Penyesuaian Kolom Dataset	45
4.1.2.	Penyesuaian Label Target	46
4.1.3.	Penggabungan Dataset.....	48
4.1.4.	Penanganan Nilai Tidak Konsisten.....	49
4.1.5.	Penyesuaian Tipe Data	50
4.1.6.	Penanganan Missing Values	51
4.1.7.	Normalisasi Data	53
4.2.	Pembagian Data.....	54
4.3.	Balancing Data	54
4.4.	Seleksi Fitur.....	55

4.5.	Implementasi Model.....	55
4.5.1.	Model Random Forest.....	56
4.5.2.	Model Random Forest dengan SMOTE-NN	56
4.5.3.	Model Random Forest dengan RFECV	57
4.5.4.	Model Random Forest dengan SMOTE-ENN dan RFECV	57
4.6.	Evaluasi Model	57
4.7.	Hasil Pengujian	62
4.7.1.	Perubahan Rasio Data 70:30	62
4.7.2.	Perubahan Rasio Data 75:25	64
4.7.3.	Perubahan Rasio Data 80:20	66
BAB V PENUTUP		73
5.1.	Kesimpulan	73
5.2.	Saran.....	74
DAFTAR PUSTAKA.....		75

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi SMOTE-ENN	16
Gambar 2.2 Ilustrasi Random Forest.....	18
Gambar 2.3 Confusion Matrix.....	20
Gambar 3.1 Kerangka Pemikiran	23
Gambar 3.2 Diagram Alur Penelitian	24
Gambar 3.3 Alur Penanganan Missing Values	27
Gambar 3.4 Alur normalisasi data	29
Gambar 3.5 Alur SMOTE-ENN	30
Gambar 3.6 Alur RFECV	35
Gambar 3.7 Alur Random Forest.....	36
Gambar 4.1 Dataset sebelum penyesuaian nama kolom	45
Gambar 4.2 Dataset setelah penyesuaian nama kolom.....	46
Gambar 4.3 Dataset sebelum penyesuaian label target	47
Gambar 4.4 Dataset setelah penyesuaian label target.....	48
Gambar 4.5 Dataset sebelum penanganan nilai tidak konsisten.....	49
Gambar 4.6 Dataset setelah penanganan nilai tidak konsisten.....	50
Gambar 4.7 Informasi tipe data pada dataset	50
Gambar 4.8 Dataset setelah konversi tipe data.....	51
Gambar 4.9 Jumlah missing values sebelum penanganan.....	51
Gambar 4.10 Jumlah missing values setelah penanganan	52
Gambar 4.11 Dataset setelah penanganan missing values.....	52
Gambar 4.12 Dataset setelah normalisasi data	53
Gambar 4.13 Hasil evaluasi model random forest	58
Gambar 4.14 Confusion matrix model random forest.....	59
Gambar 4.15 Hasil evaluasi model random forest dengan SMOTE-ENN.....	59
Gambar 4.16 Confusion matrix model random forest dengan SMOTE-ENN	60
Gambar 4.17 Hasil evaluasi model random forest dengan RFECV	60
Gambar 4.18 Confusion matrix model random forest dengan RFECV	61
Gambar 4.19 Hasil evaluasi model dengan SMOTE-ENN dan RFECV	61

Gambar 4.20 Confusion matrix model dengan SMOTE-ENN dan RFECV	62
Gambar 4.21 Hasil Akurasi Empat Skenario	68
Gambar 4.22 Hasil Precision Empat Skenario Model	69
Gambar 4.23 Hasil Recall Empat Skenario Model	70
Gambar 4.24 Hasil F1 Score Empat Skenario Model	71

DAFTAR TABEL

Tabel 3.1 Dataset	26
Tabel 3.2 Deskripsi Atribut Data	26
Tabel 3.3 Data Missing Values	27
Tabel 3.4 Data Missing Values NaN	27
Tabel 3.5 Dataset setelah penanganan missing values	28
Tabel 3.6 Data setelah normalisasi	30
Tabel 3.7 Jarak Euclidean	32
Tabel 3.8 Data setelah proses SMOTE	33
Tabel 3.9 Data setelah proses SMOTE-ENN	34
Tabel 3.10 Data awal model random forest	37
Tabel 3.11 Data bootstrap sampling pohon 1	37
Tabel 3.12 Data bootstrap sampling pohon 2	38
Tabel 3.13 Data bootstrap sampling pohon 3	38
Tabel 3.14 Skenario pengujian	42
Tabel 4.1 Hasil evaluasi pada rasio data 70:30	63
Tabel 4.2 Hasil evaluasi pada rasio data 75:25	64
Tabel 4.3 Hasil evaluasi pada rasio data 80:20	66