

**EVALUASI PERFORMA ALGORITMA KLASIFIKASI PENYAKIT
JANTUNG MENGGUNAKAN RANDOM FOREST (RF) DAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Fadhien Ramadhan Haniswaskito

20.11.3783

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**EVALUASI PERFORMA ALGORITMA KLASIFIKASI PENYAKIT
JANTUNG MENGGUNAKAN RANDOM FOREST (RF) DAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Fadhien Ramadhan Haniswaskito

20.11.3783

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

EVALUASI PERFORMA ALGORITMA KLASIFIKASI PENYAKIT JANTUNG
MENGUNAKAN RANDOM FOREST (RF) DAN SUPPORT VECTOR
MACHINE (SVM)

yang disusun dan diajukan oleh
Fadhien Ramadhan Haniswaskito
20.11.3783

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 30 Desember 2024

Dosen Pembimbing,


Mulia Sulistiyono, S.Kom M.Kom

NIK. 190302248

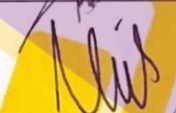
HALAMAN PENGESAHAN
SKRIPSI
EVALUASI PERFORMA ALGORITMA KLASIFIKASI PENYAKIT JANTUNG
MENGUNAKAN RANDOM FOREST (RF) DAN SUPPORT VECTOR
MACHINE (SVM)

yang disusun dan diajukan oleh

Fadhien Ramadhan Haniswaskito

20.11.3783

Telah dipertahankan di depan Dewan Penguji
pada tanggal 30 Desember 2024

Nama Penguji	Tanda Tangan
<u>Nuri Cahyono, M.Kom.</u>	
NIK. 190302278	
<u>Melany Mustika Dewi, M.Kom.</u>	
NIK. 190302455	
<u>Mulia Sulistivono, S.Kom., M.Kom.</u>	
NIK. 190302248	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 30 Desember 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

NIK. 190302096

iv

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fadhien Ramadhan Haniswaskito
NIM : 20.11.3783

Menyatakan bahwa Skripsi dengan judul berikut:

**EVALUASI PERFORMA ALGORITMA KLASIFIKASI PENYAKIT JANTUNG
MENGUNAKAN RANDOM FOREST (RF) DAN SUPPORT VECTOR
MACHINE (SVM)**

Dosen Pembimbing : Mulia Sulistiyono, S.Kom M.KOM

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 30 Desember 2024

Yang Menyatakan,



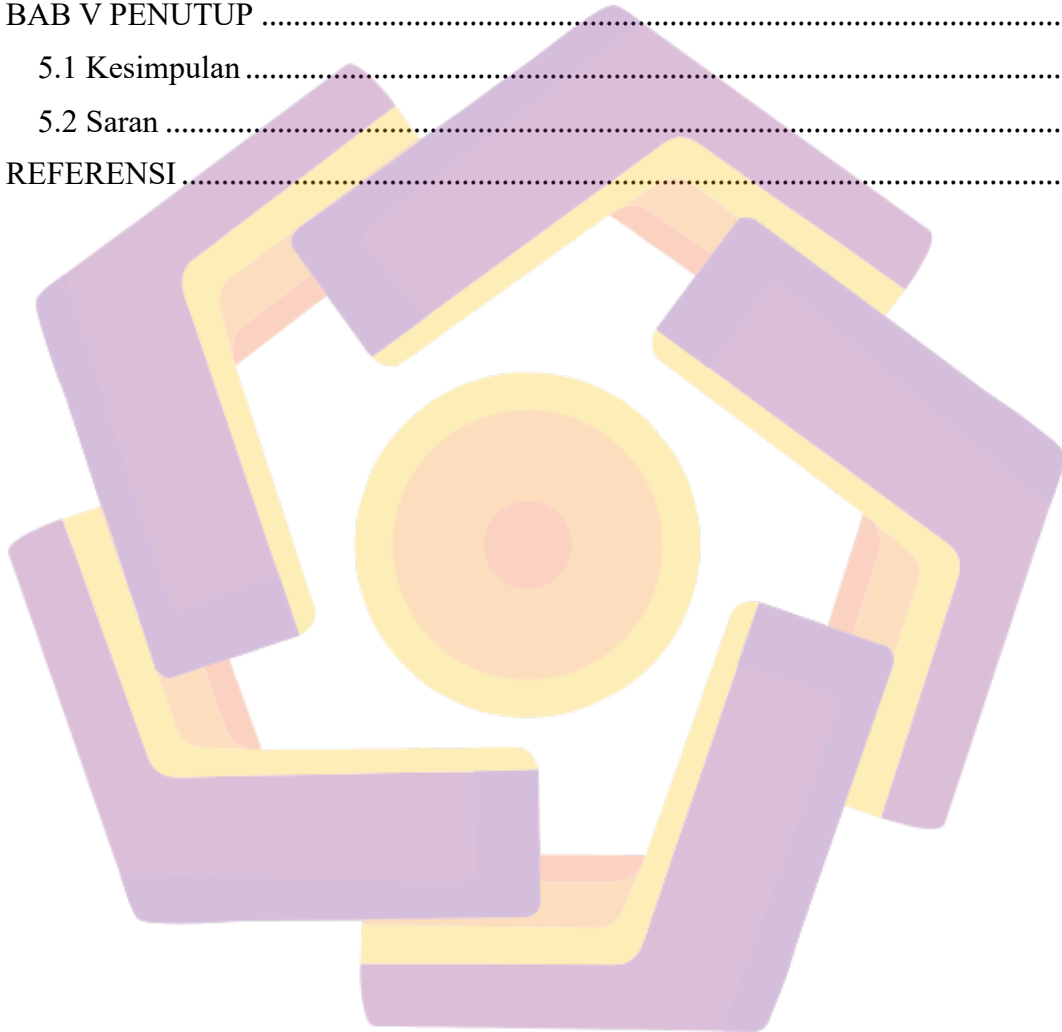
FA2A0AMX258846137

Fadhien Ramadhan Haniswaskito

DAFTAR ISI

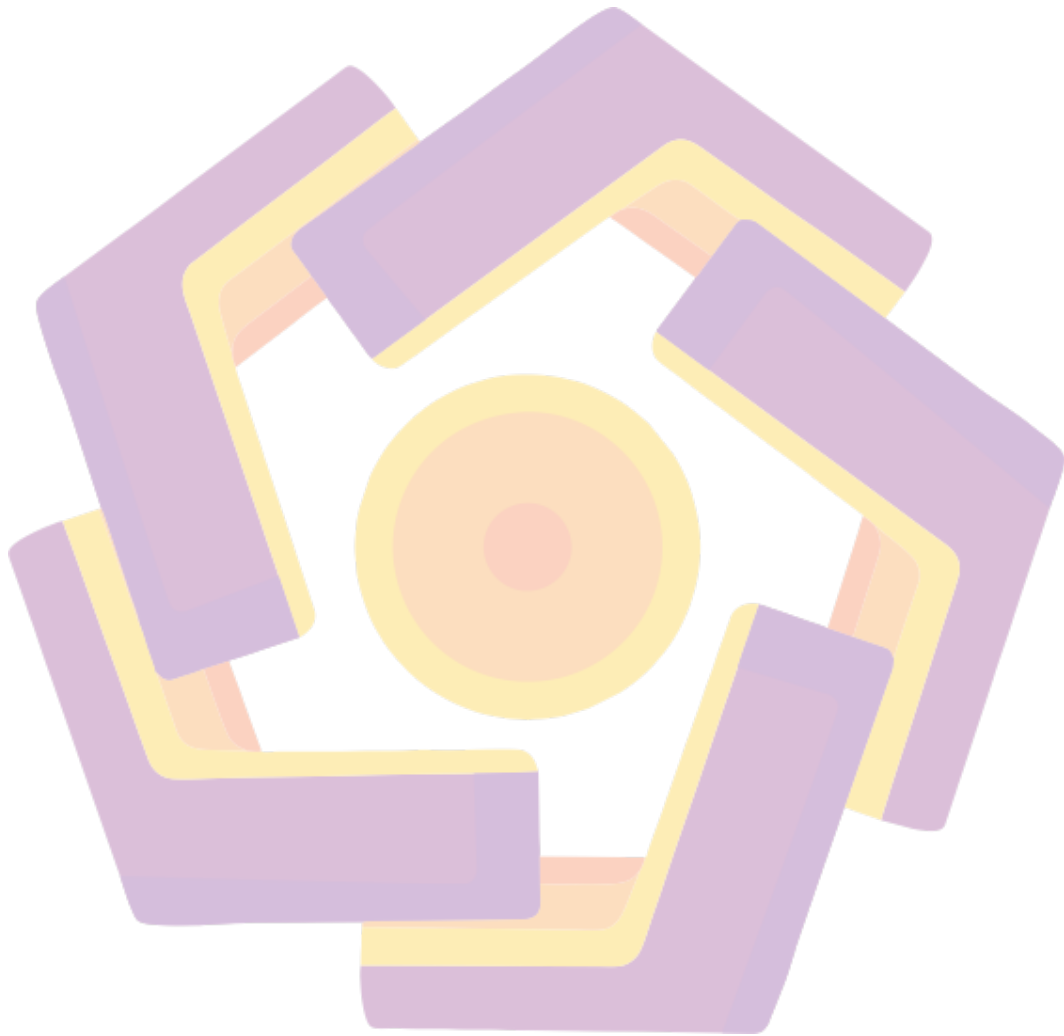
HALAMAN JUDUL	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
DAFTAR ISI.....	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
INTISARI	x
ABSTACT.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	13
2.2.1 <i>Machine Learning</i>	13
2.2.2 Evaluasi Model	18
BAB III METODE PENELITIAN	21
3.1 Alur Penelitian	21
3.1.1 Pre-Processing	21
3.1.2 Processing	23
3.1.3 Evaluation	24
3.2 Alat dan Bahan.....	24
3.2.1 Bahan Penelitian	24
3.2.2 Alat Penelitian.....	25
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Pengumpulan Data.....	26
4.2 Persiapan Data	26

4.3 Eksplorasi Data	28
4.4 Pemisahan data dan Pelatihan Model	30
4.5 Evaluasi Hasil	32
4.6 Penerapan SMOTE	35
4.7 Evaluasi Hasil Setelah SMOTE	36
4.8 Perbandingan Hasil Sebelum dan Setelah SMOTE	39
4.9 Prediksi	41
BAB V PENUTUP	43
5.1 Kesimpulan	43
5.2 Saran	44
REFERENSI	45



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 2. 2 Confusion Matrix	19
Tabel 3. 1 Penjelasan dari setiap kolom.....	24
Tabel 4. 1 Perbandingan hasil evaluasi	40



DAFTAR GAMBAR

Gambar 2. 1 Supervised Learning	13
Gambar 2. 2 Unsupervised Learning	14
Gambar 2. 3 Semi-Supervised Learning	15
Gambar 2. 4 Reinforcement Learning	15
Gambar 2. 5 Random Forest	16
Gambar 2. 6 SVM Linear	17
Gambar 2. 7 SVM RBF	17
Gambar 2. 8 SMOTE	18
Gambar 3. 1 Flowchart	21
Gambar 4. 1 Menghapus kolom yang tidak relevan	26
Gambar 4. 2 Transformasi data manual	27
Gambar 4. 3 Mengisi data kosong menggunakan 'mean'	27
Gambar 4. 4 Mengubah nama kolom agar seragam	27
Gambar 4. 5 Penggabungan ketiga dataset	28
Gambar 4. 6 Pemetaan kolom target 'num'	28
Gambar 4. 7 Deskripsi dari setiap kolom	28
Gambar 4. 8 Pengisian data '0'	29
Gambar 4. 9 Jumlah data dari setiap kolom	29
Gambar 4. 10 Korelasi antar variabel	30
Gambar 4. 11 Pemisahan data latih dan data uji	31
Gambar 4. 12 Distribusi kelas sebelum SMOTE	31
Gambar 4. 13 Pelatihan model Random Forest	32
Gambar 4. 14 Pelatihan model SVM Linear	32
Gambar 4. 15 Pelatihan model SVM RBF	32
Gambar 4. 16 Hasil Evaluasi RF sebelum SMOTE	33
Gambar 4. 17 Hasil Evaluasi RF sebelum SMOTE	34
Gambar 4. 18 Hasil Evaluasi RF sebelum SMOTE	35
Gambar 4. 19 Penerapan SMOTE	36
Gambar 4. 20 Hasil Evaluasi RF setelah SMOTE	37
Gambar 4. 21 Hasil Evaluasi SVM Linear setelah SMOTE	38
Gambar 4. 22 Hasil Evaluasi SVM RBF setelah SMOTE	39
Gambar 4. 23 Hasil Prediksi	41

INTISARI

Penyakit jantung merupakan salah satu masalah kesehatan masyarakat utama di dunia dan menjadi penyebab kematian nomor satu secara global, dengan lebih dari 17,9 juta kematian setiap tahun. Di Indonesia, penyakit jantung juga merupakan penyebab kematian utama dengan tren peningkatan prevalensi dari 0,5% pada tahun 2013 menjadi 1,5% pada tahun 2018. Pentingnya deteksi dini dan penanganan yang efektif menuntut pengembangan sistem prediksi berbasis teknologi untuk mendukung diagnosis medis.

Dalam penelitian ini, dua algoritma pembelajaran mesin, yaitu Random Forest (RF) dan Support Vector Machine (SVM), dievaluasi performanya dalam klasifikasi penyakit jantung. Dataset yang digunakan meliputi UCI Heart Disease Dataset, Heart Attack Analysis & Prediction Dataset, dan Heart Disease Dataset. Teknik oversampling Synthetic Minority Oversampling Technique (SMOTE) diterapkan untuk menangani ketidakseimbangan data.

Hasil penelitian menunjukkan bahwa algoritma RF memiliki performa lebih baik dibandingkan SVM dalam berbagai metrik evaluasi seperti ROC-AUC, precision, dan recall. Penerapan SMOTE secara signifikan meningkatkan kemampuan model dalam mendeteksi kelas minoritas, terutama pada algoritma RF yang menunjukkan peningkatan recall untuk kelas positif. RF menunjukkan fleksibilitas yang lebih baik terhadap data tidak seimbang setelah penerapan SMOTE, dengan akurasi sebesar 85% dan ROC-AUC sebesar 0,85. Oleh karena itu, RF direkomendasikan untuk digunakan dalam pengembangan sistem prediksi penyakit jantung.

Kata Kunci: Penyakit jantung, Random Forest, Support Vector Machine, SMOTE, Klasifikasi.

ABSTACT

Heart disease is one of the major global public health issues and the leading cause of death worldwide, accounting for over 17.9 million deaths annually. In Indonesia, heart disease also ranks as the primary cause of death, with prevalence increasing from 0.5% in 2013 to 1.5% in 2018. The importance of early detection and effective management necessitates the development of technology-based predictive systems to support medical diagnoses.

This study evaluates the performance of two machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), in heart disease classification. The datasets used include the UCI Heart Disease Dataset, Heart Attack Analysis & Prediction Dataset, and Heart Disease Dataset. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address data imbalance.

The results indicate that the RF algorithm outperforms SVM across various evaluation metrics, including ROC-AUC, precision, and recall. The application of SMOTE significantly improved the model's ability to detect the minority class, particularly for the RF algorithm, which demonstrated an increase in recall for the positive class. RF showed better flexibility in handling imbalanced data after SMOTE application, achieving an accuracy of 85% and a ROC-AUC score of 0.85. Therefore, RF is recommended for use in the development of heart disease prediction systems.

Keywords: Heart Disease, Random Forest, Support Vector Machine, SMOTE, Classification.