

SKRIPSI/TUGAS AKHIR
IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI
PENYAKIT JANTUNG DENGAN MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)



MUHAMMAD HILMY KHAIRY

NIM : 215610026

PROGRAM STUDI SISTEM INFORMASI

PROGRAM SARJANA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

YOGYAKARTA

2024

**TUGAS AKHIR
SKEMA SKRIPSI**

**IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI
PENYAKIT JANTUNG DENGAN MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)**

Diajukan sebagai salah satu syarat untuk menyelesaikan studi pada

Program Sarjana

Program Studi Sistem Informasi

Fakultas Teknologi Informasi

Universitas Teknologi Digital Indonesia



Disusun Oleh

MUHAMMAD HILMY KHAIRY

NIM : 215610026

PROGRAM STUDI SISTEM INFORMASI

PROGRAM SARJANA

FAKULTAS TEKNOLOGI INFORMASI

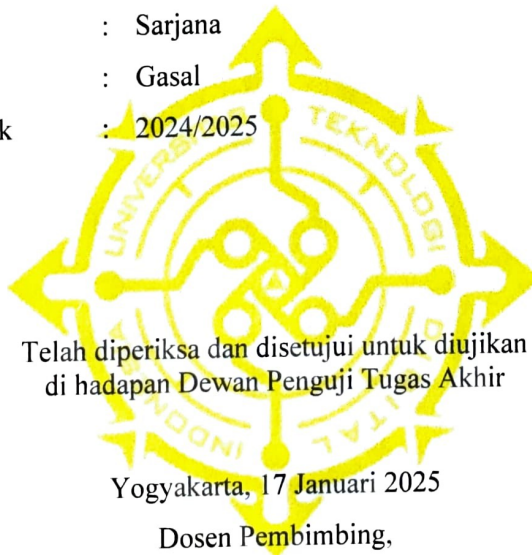
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN TUGAS AKHIR

Judul : Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Dengan Menggunakan Metode Support Vector Machine (Svm)
Nama : Muhammad Hilmy Khairy
NIM : 215610026
Program Studi : Sistem Informasi
Program : Sarjana
Semester : Gasal
Tahun Akademik : 2024/2025



Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Tugas Akhir

Yogyakarta, 17 Januari 2025

Dosen Pembimbing,

A handwritten signature in black ink, appearing to be "Robby Cokro Buwono".

Robby Cokro Buwono, S.Kom., M.Kom.
NIDN : 0529128201

HALAMAN PENGESAHAN

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI PENYAKIT JANTUNG DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Telah dipertahankan di depan Dewan Penguji dan dinyatakan diterima untuk memenuhi sebagian persyaratan guna memperoleh

Gelar Sarjana Komputer
Program Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia

Yogyakarta, 23 Januari 2025

Dewan Penguji

1. Ir. M. Guntara, M.T.
2. Robby Cokro Buwono, S.Kom., M.Kom.
3. Sumiyatun, S.Kom., M.Cs.

NIDN

0509066101
0529128201
0515048402

Tandatangan




Mengetahui

Ketua Program Studi Sistem Informasi

Deborah Kurniawati, S.Kom., M.Cs

NPP : 05111073



PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 17 Januari 2025



Muhammad Hilmy Khairy

NIM : 215610026

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah Subhanahu wa Ta'ala atas limpahan rahmat dan nikmat-Nya yang begitu besar, sehingga penulis dapat menyelesaikan skripsi ini. Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar Sarjana Komputer pada Program Studi Sistem Informasi, Universitas Teknologi Digital Indonesia.

Penulis menyadari sepenuhnya bahwa penyelesaian skripsi ini tidak akan terwujud tanpa bantuan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah berkontribusi, di antaranya:

1. Allah SWT, Tuhan yang Maha Pengasih lagi Maha Penyayang yang telah memberikan nikmat yang tak kunjung henti kepada penulis serta diberikan kelancaran dalam menyelesaikan skripsi.
2. Bapak Robby Cokro Buwono, S.Kom., M.Kom., selaku Dosen Pembimbing yang senantiasa membimbing, mengarahkan, dan memotivasi penulis selama penyusunan skripsi.
3. Sri Redjeki, S. Si., M. Kom., Ph. D., selaku Rektor Universitas Teknologi Indonesia
4. Dr. Bambang Purnomosidi DP, S. E. Akt., S. Kom., MMSI., Selau Dekan Fakultas Teknologi Informasi
5. Orang tua dan keluarga yang selalu mendoakan dan mendukung Serta memberikan motivasi.
6. Teman – teman seperjuangan Sistem Informasi Angkatan 2021 yang senantiasa memberikan dukungan semangat satu sama lain.
7. Serta semua pihak yang telah membantu penyelesaian skripsi ini yang tidak dapat disebut namanya satu per satu.

Penulis menyadari bahwa skripsi ini mungkin masih memiliki kekurangan dan kesalahan. Oleh karena itu, penulis memohon maaf atas segala kekurangan maupun kesalahan, baik dalam aspek keilmuan maupun penulisan

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR	v
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
DAFTAR LISTING	xi
INTISARI.....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	14
2.2.1 Penyakit jantung.....	14
2.2.2 Data mining.....	14
2.2.3 Support vector machine	16
2.2.4 Confusion matrix	19
2.2.5 Python	20
BAB III METODE PENELITIAN.....	22
3.1 Alur Penelitian.....	22
3.1.1 Pengumpulan Data	22
3.1.2 Pemahaman Data	22
3.1.3 Split Data	23
3.1.4 Pre-processing Data	23
3.1.5 Pemodelan.....	23
3.1.6 Evaluasi Model	23
BAB IV IMPLEMENTASI DAN PEMBAHASAN	24
4.1 Pengumpulan Data	24
4.2 Analisis dan Visualisasi data.....	26
4.2.1 Distribusi Variabel Target.....	26
4.2.2 Distribusi Variabel Usia.....	27

4.4.3 Missing Value	27
4.3 Pemrosesan Awal.....	29
4.3.1 Pembagian Data	29
4.3.1 Normalisasi	29
4.4 Implementasi dan Evaluasi Model	30
4.4.1 Hyperparameter Tuning	30
4.4.2 Evaluasi Model	31
4.5 Perhitungan manual.....	34
4.6 Pembahasan	37
BAB V KESIMPULAN DAN SARAN.....	39
5.1 Kesimpulan.....	39
5.2 Saran.....	40
BAGIAN AKHIR.....	41
Daftar Pustaka	41
Lampiran	45

DAFTAR TABEL

Tabel 2.1 Review Penelitian Terkait.....	8
Tabel 2.2 Confusin Matrix	19
Tabel 4.1 Deskripsi Atribut Dataset.....	24
Tabel 4.2 Deskripsi Penomoran Attribute.....	25
Tabel 4.3 Tabel Confusion Matrix	31
Tabel 4.4 Sampel data	34

DAFTAR GAMBAR

Gambar 2.1 Hyperplane Terbaik Yang Memisahkan Kedua Class Negatif Dan Positif	17
Gambar 3.1 Alur penelitian.....	22
Gambar 4.1 Distribusi Variabel Target.....	26
Gambar 4.2 Distribusi usia.....	27
Gambar 4.3 Hasil pemeriksaan missing value	28
Gambar 4.4 Pembagian data	29
Gambar 4.5 Confusion Matrix	32
Gambar 4.6 Hasil Klasifikasi	32
Gambar 4.7 Kurva ROC.....	33
Gambar 4.8 Hasil pengujian.....	37

DAFTAR LISTING

Listing 4.1 Potongan skrip untuk memeriksa missing value.....	28
Listing 4.2 Potongan skrip untuk normaliasi dataset	29
Listing 4.3 Potongan skrip untuk pembuatan model SVM	30
Listing 4.4 Kode skrip untuk menguji data dengan model svm.....	37

INTISARI

Penyakit jantung merupakan salah satu penyebab utama kematian di Indonesia, dengan prevalensi yang terus meningkat dari 0,5% pada 2013 menjadi 1,5% pada 2018 berdasarkan data Riset Kesehatan Dasar (Riskesdas). Kurangnya kesadaran masyarakat terhadap gejala awal penyakit ini sering menyebabkan keterlambatan diagnosis, yang berpotensi menimbulkan komplikasi serius.

Penelitian ini bertujuan untuk mengukur akurasi dalam implementasikan metode *Support Vector Machine* (SVM) untuk memprediksi penyakit jantung berdasarkan faktor-faktor risiko seperti jenis sakit dada (*chest pain*), tekanan darah (*tresbps*), kolesterol (*chol*), hasil tes EKG (*restacg*), denyut jantung (*thalach*), dan kadar gula darah puasa (*fasting blood sugar*, FBS).

Model SVM yang dikembangkan dalam penelitian ini menunjukkan akurasi keseluruhan sebesar 90% dengan nilai AUC sebesar 0,95, yang mencerminkan kemampuan sangat baik dalam membedakan individu yang berisiko dan tidak berisiko.

Kata Kunci : *Kernel RBF, Prediksi Penyakit Jantung, Support Vector Machine,*

ABSTRACT

Heart disease is one of the leading causes of death in Indonesia, with a prevalence that has increased from 0.5% in 2013 to 1.5% in 2018, according to data from the Riset Kesehatan Dasar (Riskesdas). The lack of public awareness of the early symptoms of this disease often leads to delayed diagnosis, which has the potential to cause serious complications.

This study aims to evaluate the accuracy of implementing the Support Vector Machine (SVM) method to predict heart disease based on risk factors such as chest pain type, blood pressure (tresbps), cholesterol level (chol), electrocardiogram (ECG) test results (restacg), heart rate (thalach), and fasting blood sugar (FBS).

The SVM model developed in this study demonstrated an overall accuracy of 90% with an AUC value of 0.95, reflecting excellent ability to distinguish between at-risk and non-risk individuals.

Keywords : *Heart Disease prediction, RBF Kernel, Support Vector Machine*