

**COMPARATIVE ANALYSIS OF RANDOM FOREST, SVM,
AND NAIVE BAYES FOR CARDIOVASCULAR DISEASE
PREDICTION**

LAPORAN NON-REGULER

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



Disusun oleh :

WINDY ALDORA RAYADHANI

22.11.4566

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**COMPARATIVE ANALYSIS OF RANDOM FOREST, SVM,
AND NAIVE BAYES FOR CARDIOVASCULAR DISEASE
PREDICTION**

LAPORAN NON-REGULER

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



Disusun oleh :

WINDY ALDORA RAYADHANI

22.11.4566

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**COMPARATIVE ANALYSIS OF RANDOM FOREST, SVM,
AND NAIVE BAYES FOR CARDIOVASCULAR DISEASE
PREDICTION**

yang disusun dan diajukan oleh
Windy Aldora Rayadhani

22.11.4566

telah disetujui oleh Dosen Pembimbing,
pada tanggal 15 Desember 2025

Dosen Pembimbing,



Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

HALAMAN PENGESAHAN

JALUR NON-REGULER

**COMPARATIVE ANALYSIS OF RANDOM FOREST, SVM,
AND NAIVE BAYES FOR CARDIOVASCULAR DISEASE
PREDICTION**

yang disusun dan diajukan oleh

Windy Aldora Rayadhani
22.11.4566

Telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192



Theopilus Bayu Sasoneko, S.Kom., M.Eng.
NIK. 190302375



Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Laporan ini telah diterima sebagai salah satu
persyaratan untuk memperoleh gelar Sarjana Komputer
Tanggal 15 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Windy Aldora Rayudhani

NIM : 22.11.4566

Menyatakan babwis Laporan dengan judul berikut:

Comparative Analysis of Random Forest, SVM, and Naive Bayes for Cardiovascular Disease Prediction

Dosen Pembimbing : Majid Rahardi, S.Kom, M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diujikan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali seperti 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 ketidak-benaran 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,

Yang Menyatakan, 15 Desember 2025



Windy Aldora Rayudhani

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, kebijaksanaan, dan kekuatan yang telah mengiringi setiap langkah dalam proses penyusunan laporan ini, yang merupakan hasil dari penelitian sekaligus perjalanan pembelajaran dan tanggung jawab akademik. Penulis juga menyampaikan terima kasih yang tulus atas bimbingan, dukungan, serta kepercayaan dari berbagai pihak yang telah membantu hingga laporan ini dapat terselesaikan dengan baik, antara lain kepada:

1. Bapak, Ibu, serta Mbak dan Mas tercinta, atas doa, kasih sayang, dan semangat yang tiada henti.
2. Prof. Dr. Muhammad Siyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta.
3. Prof. Dr. Kusri, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Efi Pujastuti, M.Kom, selaku Ketua Program Studi Informatika.
5. Majid Rahardi, S.Kom., M.Eng, selaku Dosen Pembimbing atas arahan dan motivasinya selama proses penyusunan laporan.
6. Tim Dosen Pengajar, atas kritik dan saran konstruktif yang membangun penelitian ini.
7. Vina, Dini, Sheila, dan Fauzan, teman-teman dekat yang selalu memberi semangat dan keceriaan.

Penulis menyadari laporan ini masih jauh dari sempurna, namun semoga dapat memberikan manfaat dan menjadi kontribusi kecil bagi pengembangan ilmu pengetahuan.

Yogyakarta, 12 Desember 2025

Windy Aldora Rayadhani

DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan Keaslian Karya.....	iv
Kata Pengantar	v
Daftar Isi.....	vi
Daftar Tabel.....	viii
Daftar Gambar.....	ix
Daftar Lampiran.....	x
Daftar Lambang dan Singkatan.....	xi
Daftar Istilah.....	xii
Intisari.....	xiii
<i>Abstract</i>	xiv
1.1 Gambaran Umum	1
1.2 Rumusan Masalah	1
1.3 Batasan Masalah.....	2
1.4 Tujuan.....	2
BAB II TINJAUAN PUSTAKA	4
2.1. Studi Literatur	4
2.2. Landasan Teori.....	5
2.2.1 Penyakit Kardiovaskular (PKV)	5
2.2.2 Data Mining dan Knowledge Discovery in Database (KDD).....	6
2.2.3 Machine Learning	8
2.2.4 Algoritma Klasifikasi dalam Machine Learning.....	11
2.2.5 Pre-Processing Data	12
2.2.6 Hyperparameter Tuning	13
2.2.7 Evaluasi Metode.....	14
2.2.8 Paired Test.....	15
2.2.9 Analisis Feature Importance (khusus Random Forest).....	16
BAB III METODE PENELITIAN	17
3.1 Jenis Penelitian.....	17

3.2	Objek Penelitian	17
3.3	Alur Penelitian.....	19
3.4	Instrumen Penelitian.....	20
3.5	Split Data.....	20
3.6	Hyperparameter Tuning dan Pelatihan Model	21
3.7	Uji Statistik (Paired Test).....	22
3.8	Analisis Feature Importance (Random Forest)	22
BAB IV PEMBAHASAN.....		24
4.1	Gambaran Umum Dataset.....	24
4.2	Pre-Processing Data	25
4.3	Split Data.....	26
4.4	Hyperparameter Tuning GridSearchCV.....	27
4.5	Hasil Pengujian Model Random Forest	28
4.6	Hasil Pengujian Model SVM	31
4.7	Hasil Pengujian Model Naïve Bayes.....	34
4.1	Perbandingan Antar Metode.....	35
4.2	Uji Signifikansi (Paired Test).....	36
4.3	Analisis Feature Importance (Khusus Random Forest)	37
BAB V Kesimpulan		39
5.1	Kesimpulan.....	39
5.2	Saran.....	40
Referensi		41
Lampiran		43
Curriculum Vitae.....		43
Lampiran Dan Bukti Pendukung.....		45
a.	Letter of Acceptance (LOA).....	45
b.	Lembar Review	46
c.	Bukti Terbit/Terindex	49
d.	Bukti Pembayaran	50

DAFTAR TABEL

Tabel 3.1.1 Dataset Deskripsi	18
Tabel 4.2.1 Rata-rata dan standar deviasi fitur sebelum dan sesudah normalisasi	26
Tabel 4.3.1 Hasil Pembagian Data	26
Tabel 4.5.1 Hasil Evaluasi Model Random Forest Baseline	28
Tabel 4.5.2 Hasil Evaluasi Model Random Forest	30
Tabel 4.5.3 Perbandingan performa Random Forest baseline	30
Tabel 4.6.1 Hasil Evaluasi Model SVM Baseline	31
Tabel 4.6.2 Hasil Evaluasi Model SVM Setelah Tuning Hyperparameter	33
Tabel 4.6.3 Perbandingan performa SVM baseline dan setelah hyperparameter tuning	33
Tabel 4.7.1 Hasil Evaluasi Model Naïve Bayes Setelah Tuning Hyperparameter	35
Tabel 4.8.1 Perbandingan Antar Metode	35
Tabel 4.9.1 Perbandingan Kinerja Model	36

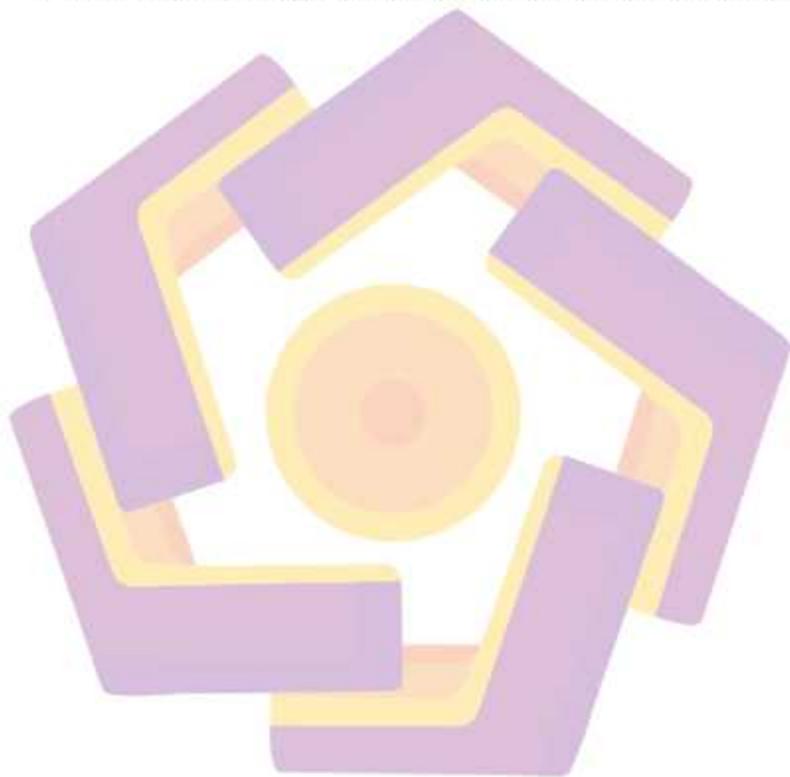


DAFTAR GAMBAR

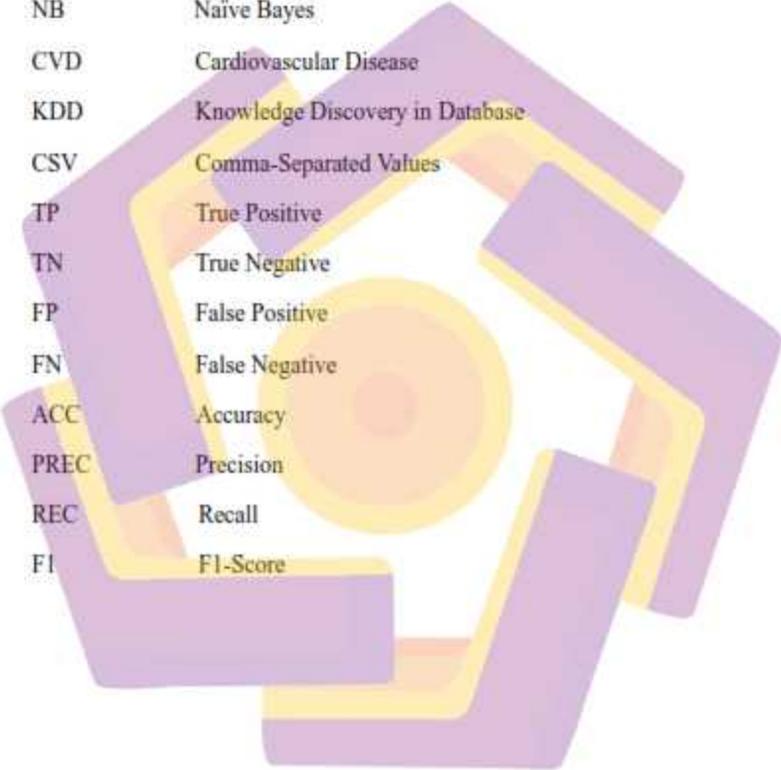
Gambar 2.2.7.1 Contoh simulasi CrossValidation	14
Gambar 3.3.1 Alur penelitian 1	20
Gambar 4.1.1 Sebagian data pada Dataset	25
Gambar 4.2.1 Hasil pengecekan terdapat missing data atau tidak	25
Gambar 4.4.1 Hyperparameter Tuning	28
Gambar 4.5.1 Confusion matrix Random Forest Baseline	28
Gambar 4.5. 2 Kurva ROC Random Forest Baseline	29
Gambar 4.5.3 Confusion matrix Random Forest Setelah Tuning Hyperparameter	29
Gambar 4.5.4 Kurva ROC Random Forest Setelah Tuning Hyperparameter	30
Gambar 4.6.1 Confusion matrix SVM Baseline	31
Gambar 4.6.2 Kurva ROC SVM Baseline	32
Gambar 4.6.3 Confusion matrix SVM Setelah Tuning Hyperparameter	33
Gambar 4.6.4 Kurva ROC SVM Setelah Tuning Hyperparameter	33
Gambar 4.7.1 Confusion matrix Naïve Bayes Setelah Tuning Hyperparameter ..	34
Gambar 4.7.2 Kurva ROC Naïve Bayes Setelah Tuning Hyperparameter	35
Gambar 4.10.1 Feature importance	38

DAFTAR LAMPIRAN

Lampiran 1. Curriculum Vitae	43
Lampiran 2 Letter of Acceptance.....	45
Lampiran 3. Lembar Review.....	46
Lampiran 4. Bukti Terindex.....	49
Lampiran 5. Bukti Terbit.....	49
Lampiran 6. Bukti Pembayaran.....	50



DAFTAR LAMBANG DAN SINGKATAN



ML	Machine Learning
RF	Random Forest
SVM	Support Vector Machine
NB	Naïve Bayes
CVD	Cardiovascular Disease
KDD	Knowledge Discovery in Database
CSV	Comma-Separated Values
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ACC	Accuracy
PREC	Precision
REC	Recall
F1	F1-Score

DAFTAR ISTILAH

<i>Machine Learning</i>	Cabang dari kecerdasan buatan (AI) yang memungkinkan komputer belajar dari data dan membuat prediksi tanpa pemrograman eksplisit.
<i>Random Forest</i>	Algoritma ensemble berbasis pohon keputusan yang menggabungkan banyak decision tree untuk meningkatkan akurasi dan mengurangi overfitting.
<i>Support Vector Machine</i>	Algoritma klasifikasi yang bekerja dengan mencari batas pemisah terbaik (hyperplane) antara dua kelas data.
<i>Naïve Bayes</i>	Algoritma probabilistik berbasis teorema Bayes dengan asumsi independensi antar fitur.
<i>Cardiovascular Disease</i>	Penyakit yang berkaitan dengan jantung dan pembuluh darah, seperti penyakit jantung koroner dan hipertensi.
<i>Dataset</i>	Kumpulan data yang digunakan untuk melatih dan menguji model machine learning.
<i>Feature</i>	Atribut atau variabel dalam dataset yang digunakan sebagai masukan model prediksi.
<i>Label / Target</i>	Nilai keluaran atau kelas yang diprediksi oleh model, seperti "memiliki penyakit" atau "tidak memiliki penyakit."
<i>Preprocessing</i>	Tahap pembersihan dan persiapan data sebelum digunakan untuk pelatihan model.
<i>Hyperparameter Tuning</i>	Proses penyesuaian parameter eksternal algoritma untuk memperoleh performa terbaik.
<i>Accuracy</i>	Ukuran yang menunjukkan persentase prediksi benar terhadap seluruh data uji.

INTISARI

Penyakit kardiovaskular merupakan salah satu penyebab utama kematian di dunia dan menjadi tantangan besar dalam bidang kesehatan masyarakat. Kasus yang terus meningkat dapat berdampak serius pada kualitas hidup dan angka harapan hidup masyarakat. Oleh karena itu, diperlukan sistem yang mampu melakukan deteksi dini secara akurat untuk mengurangi risiko fatal pada pasien berisiko tinggi. Dalam penelitian ini, dilakukan perbandingan performa tiga algoritma *machine learning*, yaitu Random Forest, Support Vector Machine (SVM), dan Naïve Bayes, dalam memprediksi risiko penyakit kardiovaskular. Dataset yang digunakan adalah *Mendeley Cardiovascular Disease Dataset* yang berisi 1.000 data pasien dengan 14 atribut klinis. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pembagian data latih dan uji, pelatihan model, evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, serta pengujian statistik menggunakan *paired t-test* untuk melihat signifikansi perbedaan performa antar model. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi 99%, recall 100%, presisi 98%, dan F1-score 99%. Model SVM dan Naïve Bayes masing-masing mencapai akurasi 98% dan 94,5%. Faktor usia, tekanan darah istirahat, dan kadar kolesterol terbukti paling berpengaruh terhadap risiko penyakit kardiovaskular. Penelitian ini memberikan kontribusi pada pengembangan sistem pendukung keputusan klinis (*Clinical Decision Support System*) yang dapat dimanfaatkan oleh tenaga medis untuk meningkatkan efektivitas deteksi dini penyakit jantung.

Kata kunci: Penyakit Kardiovaskular, Random Forest, SVM, Naïve Bayes, Sistem Pendukung Keputusan Klinis.

ABSTRACT

Cardiovascular disease is one of the leading causes of death worldwide; therefore, accurate early detection is essential to reduce fatal risks. This study aims to compare the performance of three machine learning algorithms that is Random Forest, Support Vector Machine (SVM), and Naïve Bayes in predicting cardiovascular disease risk using the Mendeley Cardiovascular Disease Dataset, which contains 1,000 patient records and 14 clinical attributes. The models were evaluated using accuracy, precision, recall, and F1-score metrics, and their performance differences were statistically tested using the paired t-test. The experimental results indicate that the Random Forest algorithm achieved the best performance with 99% accuracy, 100% recall, 98% precision, and an F1-score of 99%. The SVM model followed with 98% accuracy and 100% recall, while the Naïve Bayes algorithm obtained 94.5% accuracy and an F1-score of 95%. The p -value < 0.05 confirmed that the performance differences among the three models were statistically significant. From a clinical perspective, a model with high recall, such as Random Forest, is more desirable because it reduces the likelihood of false negatives, which are critical in heart disease diagnosis. The feature importance analysis also revealed that age, resting blood pressure, and cholesterol level were the most influential factors in predicting cardiovascular risk. These findings suggest that machine learning algorithms, particularly Random Forest, have strong potential to be implemented in Clinical Decision Support Systems (CDSS) for accurate and efficient early detection of cardiovascular disease.

Keyword: *Cardiovascular Disease, Random Forest, SVM, Naïve Bayes, Clinical Decision Support.*