

**PERBANDINGAN KINERJA MODEL *RANDOM FOREST*,
SUPPORT VECTOR MACHINE, DAN *LOGISTIC REGRESSION*
DALAM MEMPREDIKSI GAGAL JANTUNG PADA PASIEN
ICU**



Skripsi

Vincent Jonathan Budiman

00000055250

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**PERBANDINGAN KINERJA MODEL *RANDOM FOREST*,
SUPPORT VECTOR MACHINE, DAN *LOGISTIC REGRESSION*
DALAM MEMPREDIKSI GAGAL JANTUNG PADA PASIEN**

ICU



Skripsi

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Vincent Jonathan Budiman

00000055250

PROGRAM STUDI TEKNIK DAN INFORMATIKA

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Vincent Jonathan Budiman

Nomor Induk Mahasiswa : 00000055250

Program Studi : Sistem Informasi

Skripsi dengan judul:

Perbandingan Kinerja Model Random Forest, Support Vector Machine, dan Logistic Regression dalam Memprediksi Gagal Jantung pada Pasien Icu

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 4 Juni 2025



(Vincent Jonathan Budiman)

HALAMAN PENGESAHAN

Skripsi dengan judul

Perbandingan Kinerja Model Random Forest, Support Vector Machine, dan Logistic Regression dalam Memprediksi Gagal Jantung pada Pasien ICU

Oleh

Nama : Vincent Jonathan Budiman
NIM : 00000055250
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

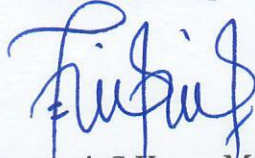
Telah diujikan pada hari Jumat, 13-Juni-2025

Pukul 10.00 s.d 12.00 dan dinyatakan

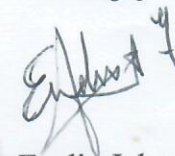
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Dr. Irmawati, S.Kom., M.M.S.I.
0805097703

Penguji


Monika Evelin Johan, S.Kom.,
M.M.S.I.
0327059501

Pembimbing


Jansen Wiratama, S.Kom., M.Kom.
0409019301

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Vincent Jonathan Budiman
NIM : 00000055250
Program Studi : Sistem Informasi
Jenjang : Strata 1 (S1)
Judul Karya Ilmiah : Perbandingan Kinerja Model Random Forest, Support Vector Machine, dan Logistic Regression dalam Memprediksi Gagal Jantung pada Pasien ICU

Menyatakan dengan sesungguhnya bahwa saya bersedia* (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 4 Juni 2025

(Vincent Jonathan Budiman)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga saya dapat menyelesaikan Skripsi yang berjudul “Perbandingan Model Random Forest, SVM, dan Logistic Regression untuk Prediksi Risiko Gagal Jantung Menggunakan Data MIMIC-IV.”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Jurusan Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.

Mengucapkan terima kasih

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Bapak Jansen Wiratama, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga penelitian ini dapat memberikan kontribusi positif dalam pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya dalam upaya deteksi dini risiko gagal jantung pada pasien ICU.

Tangerang, 4-Juni-2025



(Vincent Jonathan Budiman)

PERBANDINGAN KINERJA MODEL RANDOM FOREST, SUPPORT VECTOR MACHINE, DAN LOGISTIC REGRESSION DALAM MEMPREDIKSI GAGAL JANTUNG PADA PASIEN ICU

(Vincent Jonathan Budiman)

ABSTRAK

Penyakit gagal jantung merupakan salah satu penyebab mortalitas tertinggi di kalangan pasien *Intensive Care Unit* (ICU), sehingga menuntut adanya metode deteksi dini yang efektif untuk mencegah komplikasi fatal. Kemajuan teknologi *machine learning* menawarkan potensi besar untuk membantu proses prediksi dan pengambilan keputusan klinis secara lebih cepat dan akurat. Penelitian ini bertujuan untuk membandingkan secara komprehensif kinerja tiga algoritma klasifikasi—*Random Forest*, *Support Vector Machine* (SVM), dan *Logistic Regression*—dalam memprediksi risiko gagal jantung pada pasien ICU dengan memanfaatkan dataset klinis komprehensif MIMIC-IV.

Mengadopsi kerangka kerja CRISP-DM, penelitian ini melalui proses yang sistematis mulai dari persiapan data, pemodelan, hingga evaluasi. Hasil evaluasi menunjukkan adanya *trade-off* yang signifikan antar model: *Random Forest* menunjukkan performa diskriminatif terbaik secara keseluruhan dengan akurasi 89,47% dan ROC-AUC 0,9401. Sebaliknya, SVM, meskipun dengan akurasi lebih rendah (81,58%), memiliki nilai *recall* tertinggi sebesar 66,67%. Analisis lebih lanjut menggunakan *feature importance* dari *Random Forest* mengidentifikasi bahwa detak jantung (*heart rate*), saturasi oksigen (O2 *saturation*), dan usia (*anchor age*) adalah tiga faktor prediktor yang paling berpengaruh secara berurutan.

Dengan mempertimbangkan konteks penerapan di lingkungan klinis, di mana kegagalan mendeteksi kondisi kritis (*false negative*) membawa risiko yang jauh lebih besar, model SVM dinilai sebagai yang paling efektif dan tepat. Kemampuan superior SVM dalam hal *recall* memastikan sistem dapat meminimalkan risiko terlewatnya pasien yang memerlukan intervensi medis segera, menjadikannya pilihan yang paling bertanggung jawab dan dapat diandalkan untuk implementasi nyata dalam sistem pendukung keputusan klinis di ICU.

Kata kunci: Gagal Jantung, ICU, MIMIC-IV, Pembelajaran Mesin, Prediksi

**COMPARISON OF RANDOM FOREST, SUPPORT VECTOR
MACHINE, AND LOGISTIC REGRESSION MODEL
PERFORMANCE IN PREDICTING HEART FAILURE IN ICU
PATIENTS**

(Vincent Jonathan Budiman)

ABSTRACT (English)

Heart failure is a leading cause of mortality among patients in the Intensive Care Unit (ICU), demanding effective early detection methods to prevent fatal complications. Advances in machine learning technology offer significant potential to assist in the prediction and clinical decision-making process more quickly and accurately. This study aims to comprehensively compare the performance of three classification algorithms—Random Forest, Support Vector Machine (SVM), and Logistic Regression—in predicting the risk of heart failure in ICU patients by leveraging the comprehensive MIMIC-IV clinical dataset.

Adopting the CRISP-DM framework, this research followed a systematic process from data preparation and modeling to evaluation. The evaluation results revealed a significant trade-off between the models: Random Forest demonstrated the best overall discriminative performance with an accuracy of 89.47% and a ROC-AUC of 0.9401. Conversely, SVM, despite a lower accuracy of 81.58%, achieved the highest recall value of 66.67%. Further analysis using Random Forest's feature importance identified that heart rate, O2 saturation, and anchor age are the three most influential predictors, respectively.

Considering the context of clinical application, where the failure to detect a critical condition (a false negative) carries a much greater risk, the SVM model is judged to be the most effective and appropriate. SVM's superior recall capability ensures that the system can minimize the risk of overlooking patients requiring immediate medical intervention, making it the most responsible and reliable choice for real-world implementation in a clinical decision support system in the ICU.

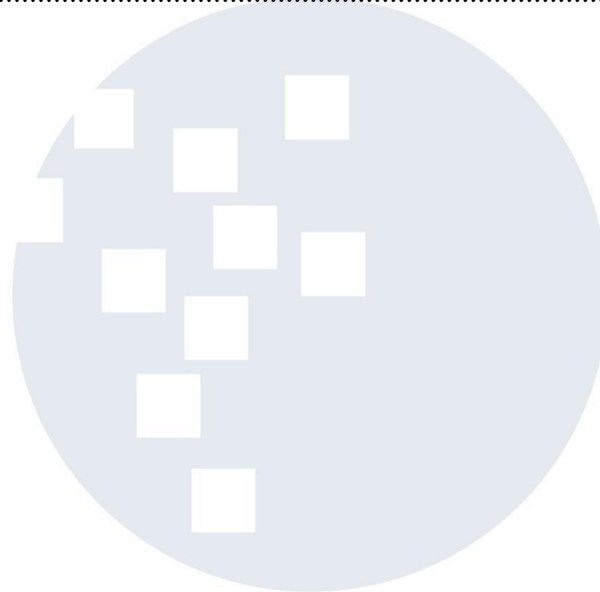
Keywords: *Heart Failure, ICU, Machine Learning, MIMIC-IV, Prediction*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan dan Manfaat Penelitian	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Penelitian Terkait.....	8
2.2 Teori Penelitian	12
2.2.1 Dataset MIMIC-IV.....	12
2.2.2 <i>Intensive Care Unit (ICU)</i>	13
2.2.3 Gagal Jantung	14
2.2.4 Detak jantung	15
2.2.5 <i>Systolic dan Diastolyc</i>	16
2.2.6 <i>Respiratory</i>	16
2.2.7 SPO2	17
2.2.8 Sodium.....	17

2.2.9	BNP	18
2.2.10	<i>Creatinine</i>	19
2.2.11	<i>Potassium</i>	20
2.3	Framework/Algoritma	20
2.3.1	<i>Logistic Regression</i>	23
2.3.2	Random Forest	24
2.3.3	Support Vector Machine	26
2.3.4	Receiver Operating Characteristic (ROC) dan Area Under the Curve (AUC)	30
2.3.5	Akurasi	31
2.3.6	Recall	32
2.4	Tools Penelitian	33
BAB III METODOLOGI PENELITIAN		35
3.1	Gambaran Umum Objek Penelitian	35
3.2	Metode Penelitian	37
3.3	Teknik Pengumpulan Data	41
3.4	Teknik Analisis Data	41
3.4.1	Variabel Penelitian	43
BAB IV ANALISIS DAN HASIL PENELITIAN		46
4.1	Tahap <i>Business Understanding</i>	46
4.2	Tahap <i>Data Understanding</i>	49
4.3	Tahap <i>Data Preparation</i>	57
4.4	Data Modelling	81
4.5	Evaluation	82
4.5.1	Random Forest	84
4.5.2	SVM	85
4.5.3	Logistic Regression	86
4.5.4	Visualisasi ROC AUC	87
4.5.5	Mencari Variabel yang paling signifikan	88
4.6	Analisis dan Diskusi	90
4.6.1	Analisis Performa Model Algoritma	91
4.6.2	Analisis Faktor Paling Berpengaruh (<i>Feature Importance</i>)	92

4.6.3 Perbandingan dengan Penelitian Terdahulu.....	93
BAB V SIMPULAN DAN SARAN.....	96
5.1 Simpulan	96
5.2 Saran	97
DAFTAR PUSTAKA.....	98
LAMPIRAN.....	106



UMN
 UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	8
Tabel 2. 2 Perbandingan Kerangka Kerja Data Mining.....	38
Tabel 4. 1 Penjelasan dataset yang digunakan	49
Tabel 4. 2 ID yang digunakan pada data Charthevents.....	52
Tabel 4. 3 Data yang digunakan pada data Labevent.....	54
Tabel 4. 4 Perbandingan penelitian ini dengan penelitian terdahulu	94



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

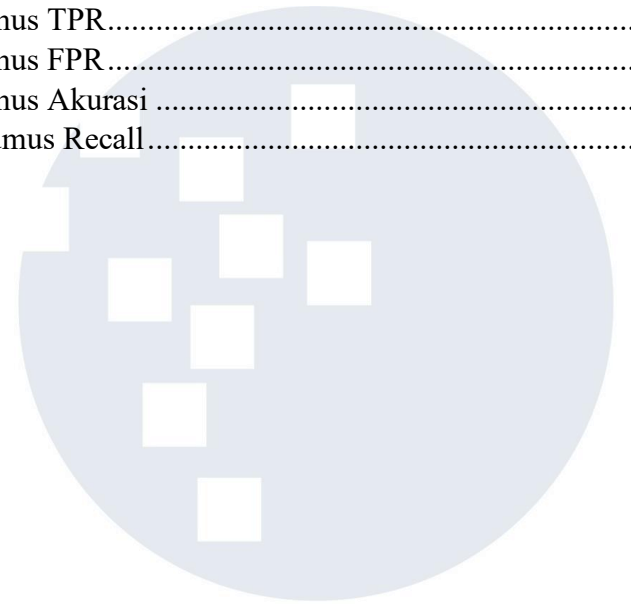
Gambar 2. 1 CRISP-DM (Cross Industry Standard Process for Data Mining) [50]	21
Gambar 2. 2 Ilustrasi Linear Regression dan Logistic Regression	23
Gambar 2. 3 Ilustrasi Random Forest	25
Gambar 2. 4 Ilustrasi Support Vector Machine	26
Gambar 3. 1 Research Freamework	41
Gambar 4.1 Syntax Import Library	48
Gambar 4. 2 Read data MIMIC IV	48
Gambar 4. 3 Visualisasi Data Patient	49
Gambar 4. 4 Info dari dataset patients	50
Gambar 4. 5 Visualisasi Data Chartevents	51
Gambar 4. 6 Info dari dataset Chartevents	51
Gambar 4. 7 Mencari ID dari setiap label yang digunakan	52
Gambar 4. 8 Visualisasi Data Labevent	53
Gambar 4. 9 Info dari data labevents	53
Gambar 4. 10 Code untuk memilih indicator	54
Gambar 4. 11 Visualisasi Data Diagnosed	55
Gambar 4. 12 Info dari data diagnoses	56
Gambar 4. 13 Visualisasi Data d_icd_diagnoses	57
Gambar 4. 14 Filtering Data Patient	58
Gambar 4. 15 Code untuk filter data chartevents sesuai dengan ID yang telah ditentukan	59
Gambar 4. 16 Code untuk filter kolom pada data chartevents	60
Gambar 4. 17 Code untuk filter data Labevent sesuai dengan ID yang telah ditentukan	60
Gambar 4. 18 Code untuk memfilter kolom data Labevent	61
Gambar 4. 19 Filter data d_icd_diagnoses	62
Gambar 4. 20 Merge data diagnosis	62
Gambar 4. 21 Penggabungan data Patient dengan df_hearts	63
Gambar 4. 22 Proses menghapus data yang duplikat pada data heartfailure_diagnose	64
Gambar 4. 23 Filtering kolom pada data heartfailure_diagnose	64
Gambar 4. 24 Membaca data diagnoses_icd(heart_failure)	65
Gambar 4. 25 Membaca data Patient	66
Gambar 4. 26 Menambahkan kolom 'heart failure' pada data 'patient'	66
Gambar 4. 27 Penggabungan data patients dengan data chartevents	67
Gambar 4. 28 Filter kolom pada data merge1	68
Gambar 4. 29 Penggabungan data patients dengan data labevents	69
Gambar 4. 30 Filtering data merge2	69
Gambar 4. 31 membersihkan data merge1 dan merge2	70
Gambar 4. 32 penggabungan data merge1 dan merge2	71

Gambar 4. 33 Membaca data Merged_Data.csv	73
Gambar 4. 34 Proses Grouping Data Berdasarkan subject_id dan itemid	74
Gambar 4. 35 Nunique dari data df_grouped.....	75
Gambar 4. 36 Proses Transformasi Data dengan Pivot Table	75
Gambar 4. 37 Proses Penggantian Nama Kolom Berdasarkan itemid.....	76
Gambar 4. 38 Proses reset index pada df.....	77
Gambar 4. 39 Tahap membuang kolom yang duplikat.....	77
Gambar 4. 40 info dari data yang digunakan	78
Gambar 4. 41 Tahap Splitting Data.....	79
Gambar 4. 42 Pipeline preprocessing setelah dilakukan split.....	80
Gambar 4. 43 Pendefinisian tiga model klasifikasi dalam satu models.	81
Gambar 4. 44 Pembuatan pipeline lengkap yang terdiri dari preprocessing, SMOTE, dan model klasifikasi.	81
Gambar 4. 45 Proses pelatihan dan prediksi model terhadap data latih dan data uji.	82
Gambar 4. 46 Code untuk memvisualisasi hasil Evaluation.....	83
Gambar 4. 47 Hasil Random Forest.....	84
Gambar 4. 48 Hasil dari SVM	85
Gambar 4. 49 Hasil dari Logistic Regression	86
Gambar 4. 50 Visualisasi ROC-AUC dari ketiga model	87
Gambar 4. 51 Code Analisis Feature Importance	88
Gambar 4. 52 Hasil dari pencarian feature_importance dari data yang digunakan	89



DAFTAR RUMUS

Rumus 2. 1 Rumus Logistic Regression	23
Rumus 2. 2 Penjelasan variabel z.....	23
Rumus 2. 3 Rumus Linear.....	27
Rumus 2. 4 Rumus Polynomial.....	28
Rumus 2. 5 Rumus RBL	29
Rumus 2. 6 Rumus Sigmoid	29
Rumus 2. 7 Rumus TPR.....	30
Rumus 2. 8 Rumus FPR.....	30
Rumus 2. 9 Rumus Akurasi	31
Rumus 2. 10 Rumus Recall.....	32



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1 Turnitin Similarity Report	106
Lampiran 2 Form Konsultasi Bimbingan.....	107
Lampiran 3 Sertifikat <i>Collaborative Institutional Training Initiative</i> (CITI PROGRAM).....	108



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA