

# Seleksi Fitur Berbasis Mutual Information untuk Optimalisasi Model Prediksi Tingkat Kematian Penderita Gagal Jantung Menggunakan Machine Learning

Elsa Sari Hayunah Nurdiniyah<sup>1\*</sup>, Aisya Nur Aulia Yusuf<sup>2</sup>, Rahardian Luthfi Prasetyo<sup>3</sup>, Rasyida Shabihah Zukron Aini<sup>4</sup>

<sup>1</sup> Universitas Jenderal Soedirman; email : [elsa.nurdiniyah@unsoed.ac.id](mailto:elsa.nurdiniyah@unsoed.ac.id)

<sup>2</sup> Universitas Jenderal Soedirman; email : [aisya.yusuf@unsoed.ac.id](mailto:aisya.yusuf@unsoed.ac.id)

<sup>3</sup> Universitas Wijayakusuma Purwokerto; email : [rahardian@unwiku.ac.id](mailto:rahardian@unwiku.ac.id)

<sup>4</sup> Institut Teknologi Sepuluh Noverber; email : [rasyida.aini@its.ac.id](mailto:rasyida.aini@its.ac.id)

Penulis Korespondensi : Elsa Sari Hayunah Nurdiniyah

**Abstract:** Heart failure is one of the leading causes of death in the world that requires an accurate and efficient prediction system to support clinical decision making. This study aims to develop a prediction model for the risk of death in patients with heart failure by optimizing feature selection using the Mutual Information (MI) approach. The main problem raised is the high complexity of clinical data with many features that are not always relevant, which can reduce the accuracy and efficiency of predictive models. The method proposed in this study involves MI-based feature selection to identify the most informative features against the target variable (patient mortality), which are then used to train various machine learning algorithms such as Random Forest, Gradient Boosting, XGBoost, and Logistic Regression. The hyperparameter tuning process is performed to optimize the performance of each model. The test results show that the Random Forest model that has been tuned using five selected features managed to achieve an accuracy of 0.99 and F1-score of 0.99, outperforming other models in terms of balance between accuracy and generalization. The results show that Mutual Information is effective in simplifying model complexity without compromising prediction performance.

**Keywords:** heart failure; mutual information; random forest; feature selection; machine learning

**Abstrak:** Gagal jantung merupakan salah satu penyebab utama kematian di dunia yang memerlukan sistem prediksi yang akurat dan efisien untuk mendukung pengambilan keputusan klinis. Penelitian ini bertujuan untuk mengembangkan model prediksi risiko kematian pada penderita gagal jantung dengan mengoptimalkan seleksi fitur menggunakan pendekatan Mutual Information (MI). Permasalahan utama yang diangkat adalah tingginya kompleksitas data klinis dengan banyak fitur yang tidak selalu relevan, sehingga dapat menurunkan akurasi dan efisiensi model prediktif. Metode yang diusulkan dalam penelitian ini melibatkan seleksi fitur berbasis MI untuk mengidentifikasi fitur paling informatif terhadap variabel target (kematian pasien), yang kemudian digunakan untuk melatih berbagai algoritma machine learning seperti Random Forest, Gradient Boosting, XGBoost, dan Logistic Regression. Proses tuning hyperparameter dilakukan untuk mengoptimalkan kinerja masing-masing model. Hasil pengujian menunjukkan bahwa model Random Forest yang telah dituning menggunakan lima fitur terpilih berhasil mencapai akurasi 0.99 dan F1-score 0.99, mengungguli model lain dalam hal keseimbangan antara akurasi dan generalisasi. Hasil penelitian ini menunjukkan bahwa Mutual Information efektif dalam menyederhanakan kompleksitas model tanpa mengorbankan performa prediksi.

**Kata kunci:** Gagal Jantung; Mutual Information; Random Forest; Seleksi Fitur; Machine Learning

Diterima: 10 Juli 2025

Direvisi: 15 Juli 2025

Diterima: 19 Juli 2025

Diterbitkan: 29 Juli 2025

Versi sekarang: 29 Juli 2025



Hak cipta: © 2025 oleh penulis.  
Diserahkan untuk kemungkinan publikasi akses terbuka berdasarkan syarat dan ketentuan lisensi Creative Commons Attribution (CC BY SA) (<https://creativecommons.org/licenses/by-sa/4.0/>)

## 1. Pendahuluan

Gagal jantung (heart failure) merupakan penyakit kardiovaskular yang sangat berbahaya dan menjadi salah satu penyebab kematian terbesar di dunia. Diperkirakan jumlah penderita gagal jantung di seluruh dunia mencapai sekitar 64 juta orang sehingga gagal jantung saat ini dipandang sebagai kondisi global yang menyerupai pandemi [1]. Diperkirakan angka ini akan terus meningkat seiring bertambahnya usia populasi secara global. Di Amerika Serikat, proyeksi terbaru menunjukkan bahwa prevalensi gagal jantung dapat meningkat hingga 46% antara tahun 2012 hingga 2030, disertai dengan lonjakan biaya layanan kesehatan yang diperkirakan mencapai 127% [2]. Di Indonesia sendiri belum tersedia data nasional terbaru tentang prevalensi gagal jantung, tetapi hasil Riset Kesehatan Dasar 2018 menyebutkan prevalensi penyakit jantung mencapai sekitar 1,5%, dengan kecenderungan peningkatan pada kelompok usia produktif [3]. Gagal jantung adalah kondisi kronis yang kompleks di mana jantung gagal memompa darah ke seluruh tubuh diantaranya disebabkan oleh diabetes, tekanan darah tinggi, atau penyakit dan gangguan jantung lainnya [4]. Jika tidak ditangani dengan cepat dan tepat, penyakit ini dapat menyebabkan penurunan kualitas hidup yang signifikan hingga kematian. Sayangnya, masih terbatas fasilitas dan sistem yang mampu memprediksi tingkat keparahan atau potensi kematian akibat gagal jantung secara cepat dan akurat, khususnya di fasilitas kesehatan tingkat dasar.

Penanganan gagal jantung menuntut kecepatan dan akurasi dalam pengambilan keputusan klinis. Salah satu pendekatan modern yang menjanjikan untuk mendukung hal ini adalah penerapan machine learning (ML) dalam prediksi risiko kematian pasien. ML memungkinkan pemodelan data medis secara otomatis dan adaptif, termasuk dalam memetakan hubungan kompleks antar variabel klinis untuk mengklasifikasikan kondisi pasien. Beberapa penelitian telah mengaplikasikan algoritma ML untuk klasifikasi penyakit gagal jantung dengan berbagai pendekatan seleksi fitur dan algoritma klasifikasi. Sebagian besar studi memanfaatkan dataset *Heart Failure Clinical Records* dari UCI dan berfokus pada peningkatan akurasi model melalui pemilihan fitur yang relevan. Penelitian oleh Chicco, dkk menunjukkan bahwa hanya dua fitur—*serum creatinine* dan *ejection fraction*—cukup untuk menghasilkan performa prediksi yang kuat menggunakan algoritma Random Forest, dengan F1-score mencapai 0.754 [5]. Sementara itu, studi oleh Oladosu, dkk membandingkan empat metode seleksi fitur statistik, yaitu *Information Gain*, *Gain Ratio*, *Symmetrical Uncertainty*, dan *Correlation*, dan menyimpulkan bahwa Random Forest memberikan hasil paling konsisten dengan akurasi 0.83 dan AUC 0.906 [6]. Pendekatan ini menekankan pentingnya pemilihan atribut yang tepat untuk meningkatkan performa klasifikasi.

Studi oleh Asif Newaz, dkk menerapkan kombinasi Balanced Random Forest dengan metode seleksi fitur *Chi-square* dan *Recursive Feature Elimination (RFE)*. Hasil yang diperoleh cukup kuat, yaitu akurasi 0.76, sensitivitas 0.80, dan AUC 0.77 untuk metode Balanced Random Forest dan *Chi-square*, yang mana menunjukkan efektivitas metode seleksi fitur *Chi-square* dalam mengatasi ketidakseimbangan data [7]. Kemudian penelitian yang dilakukan oleh Wu, dkk mencoba menggunakan skema metode seleksi fitur baru dengan mengintegrasikan antara model *cox proportional hazards* dan Random Forest diperoleh akurasi mencapai 0.754 [8]. Qadri, dkk menggunakan *transfer learning* baru untuk menseleksi dan 9 algoritma machine learning lainnya diperoleh hasil akurasi terbaik mencapai 0.975 untuk random forest [9]. Di sisi lain, penelitian oleh Mochammad Anshori, dkk menunjukkan bahwa seleksi fitur menggunakan *Information Gain* berhasil meningkatkan akurasi Logistic Regression dari 0.838 menjadi 0.856 [10].

Penelitian-penelitian tersebut menunjukkan pentingnya seleksi fitur dalam proses prediksi risiko kematian pada gagal jantung dan dengan pemilihan fitur yang baik memungkinkan model ML mencapai akurasi tinggi meski dengan jumlah fitur yang lebih sedikit [11]. Namun dari beberapa penelitian sebelumnya, metode seleksi fitur yang ada belum menunjukkan hasil yang optimum dan mayoritas studi masih bergantung pada metode seleksi fitur klasik (*Information Gain*, *Chi-Square*, korelasi), yang belum tentu menangkap pola kompleks dalam data klinis. Salah satu metode seleksi fitur lain yang belum secara khusus dikaji pada prediksi resiko kematian gagal jantung adalah *Mutual Information*. *Mutual Information (MI)* cocok untuk mengukur ketergantungan nonlinier antara fitur dan label, suatu aspek yang sangat relevan untuk data klinis di mana hubungan antar fitur bisa sangat kompleks dan tidak linier [12]. Penelitian terkini menunjukkan bahwa penggunaan MI untuk seleksi fitur secara signifikan meningkatkan akurasi model klasifikasi penyakit jantung. Ahmad, dkk misalnya,

membandingkan MI dengan ANOVA dan Chi-square untuk dataset penyakit jantung, dan menemukan MI menghasilkan akurasi tertinggi 0.82 serta meningkatkan kinerja algoritma seperti Logistic Regression dan Random Forest [13]. Selain itu, MI bersifat model-agnostik dimana MI tidak bergantung pada algoritma klasifikasi tertentu, sehingga dapat digunakan sebagai tahap filter awal sebelum diterapkan pada berbagai algoritma machine learning [14]. Keunggulan lainnya adalah stabilitas seleksi fitur MI yang tinggi, algoritma berbasis MI seperti DMIML dan HOCMIM mempertahankan konsistensi seleksi meskipun data ketika dilakukan pada subset data yang berbeda [15]. Hal ini memberikan potensi generalisasi yang baik dalam aplikasi prediksi medis.

Berdasarkan hal tersebut, dalam penelitian ini digunakan metode seleksi fitur berbasis Mutual Information sebagai langkah utama untuk mengoptimalkan performa model klasifikasi gagal jantung. Untuk membangun model klasifikasi yang robust, kami mengevaluasi beberapa algoritma ML yang umum digunakan dalam diagnosis medis: (a) Random Forest, (b) Logistic Regression, (c) XGBoost, dan (d) Gradient Boosting. Algoritma-algoritma ini telah banyak diteliti dalam konteks penyakit jantung. Misalnya, metode Random Forest sering unggul dalam berbagai studi klasifikasi jantung, sedangkan Logistic Regression menyediakan model sederhana namun seringkali efektif. XGBoost dan Gradient Boosting dipilih karena kemampuannya menangkap hubungan nonlinier kompleks melalui ensemble learning. Dengan membandingkan keempat algoritma ini, diharapkan diperoleh pemahaman mendalam tentang kekuatan dan kelemahan masing-masing dalam klasifikasi gagal jantung, serta model terbaik yang dapat diadopsi. Kontribusi makalah ini meliputi:

1. Penerapan seleksi fitur berbasis Mutual Information untuk menyeleksi variabel klinis paling relevan dalam klasifikasi gagal jantung.
2. Perbandingan komprehensif sejumlah algoritma ML (Random Forest, Logistic Regression, XGBoost, dan Gradient Boosting) dalam konteks prediksi resiko kematian gagal jantung, guna menentukan model dengan performa terbaik.

Selanjutnya pada Tinjauan Literatur berisi penjelasan terkini metode prediksi kematian pada penderita gagal jantung baik dari segi metode seleksi fitur maupun algoritma machine learning yang digunakan. Kemudian pada bab Metode akan diuraikan data yang digunakan dan metode penelitian (termasuk proses seleksi fitur Mutual Information dan skema perbandingan algoritma). Bab Hasil dan Pembahasan menyajikan hasil eksperimen serta diskusi tentang performa model yang dihasilkan. Pada bagian Perbandingan akan dibahas mengenai hasil penelitian yang sudah dilakukan dengan penelitian sebelumnya. Akhirnya, pada bagian Kesimpulan dijelaskan mengenai temuan penting, kontribusi utama, dan arah penelitian lanjutan.

## 2. Tinjauan Literatur

### 2.1. Prediksi Resiko Kematian pada Pasien Gagal Jantung

Beberapa tahun terakhir penelitian berkaitan dengan prediksi resiko kematian pasien gagal jantung sudah dilakukan. Chicco, dkk memanfaatkan Random Forest dan analisis biostatistik untuk menilai kontribusi fitur pada prediksi mortalitas pasien gagal jantung. Mereka menemukan bahwa serum creatinine dan ejection fraction sebagai dua fitur paling penting, diikuti age dan creatinine phosphokinase. Dengan hanya kedua fitur utama ini, model Random Forest memperoleh F1-score 0.754, *true negative rate* 0.855, dan akurasi 0.585. Gradient Boosting unggul pada PR AUC (0.673) dan ROC AUC (0.792), sedangkan Support Vector Machine (SVM) radial paling lemah [5].

Oladimeji, dkk kemudian mengatasi ketidakseimbangan data dengan hybrid SMOTE+ENN dan membandingkan empat filter method (Symmetrical Uncertainty, Information Gain, Gain Ratio, Correlation). Keempat metode menegaskan serum creatinine dan ejection fraction sebagai fitur paling relevan, serta smoking dan age. Setelah seleksi fitur, Random Forest mencatat akurasi 0.83, ROC AUC 0.906, dan PR AUC 0.9046 [6].

Ishaq, dkk menggunakan SMOTE untuk mengatasi ketidakseimbangan data pada data yang digunakan. Selain itu mereka juga menggunakan Random Forest pada dataset UCI untuk memperoleh fitur-fitur teratas yang berkaitan dengan hasil prediksi dan menyisakan sembilan fitur teratas. Algoritma machine learning yang mereka gunakan ada 9 yakni Decision Tree (DT), Adaptive boosting classifier (AdaBoost), Logistic Regression (LR), Stochastic Gradient

classifier (SGD), Random Forest (RF), Gradient Boosting classifier (GBM), Extra Tree Classifier (ETC), Gaussian Naive Bayes classifier (G-NB), dan Support Vector Machine (SVM). Hasil prediksi tertinggi dengan semua fitur didapatkan oleh algoritma Extra Trees Classifier (ETC) yang mencapai akurasi 0.9262 dan precision/recall/F1  $\approx$  0.93; RF dan AdaBoost sedikit di bawah. Sedangkan hasil prediksi dengan sembilan fitur terpilih tidak menurunkan performa ETC (0.9262) dan RF (0.9188) bahkan pada algoritma lain menunjukkan adanya peningkatan, membuktikan efisiensi seleksi fitur [16].

Zaman, dkk membandingkan Decision Tree, Random Forest, dan XGBoost setelah melakukan SMOTE pada data. Zaman memutuskan untuk tidak menseleksi fitur dan menggunakan seluruh fitur (12 fitur) yang dimiliki. Untuk prediksi, Zaman menerapkan Stacked Ensemble ML, dan memperoleh akurasi, precision, recall, dan F1-score 0.99, melampaui studi sebelumnya [17].

Newaz, dkk mengombinasikan Balanced Random Forest (BRF) dengan Chi-square dan RFE sebagai metode seleksi fitur. Newaz selanjutnya hanya memilih lima hingga tujuh fitur utama. BRF+Chi-square mencetak performansi terbaik dengan akurasi 0.76, sensitivitas 0.80, dan AUC 0.77 [7].

Adi, dkk membandingkan SVM, KNN, dan RF tanpa menggunakan seleksi fitur. Hasilnya diperoleh untuk algoritma RF dan SVM sama-sama mencapai akurasi 0.97, sedangkan KNN 0.93 [18].

Wu, dkk memperkenalkan empat skema seleksi fitur (“Program One”–“Four”) dengan 3–6 variabel, lalu menerapkan MLP (“Our Model”) sebagai algoritma machine learning yang digunakan. Akurasi tertinggi mencapai 0.754, mengungguli RF (0.740), GB (0.738), NB (0.696), dan ANN konvensional (0.680), menegaskan kecukupan kombinasi fitur minimalis [8].

Qadri, dkk menerapkan SMOTE untuk mengatasi imbalanced data dan transfer learning sebagai metode seleksi fitur pada 9 algoritma machine learning yang digunakan (LR, RF, SVM, Decision Tree, XGB, Naive Bayes, KNN, ETC, Gradient Boost). Transfer learning pada RF menghasilkan akurasi 0.975 dan precision/recall/F1-score 0.98 [9].

Sitanggang, dkk fokus pada tuning hyperparameter RF menggunakan Random dan Grid Search; Grid Search menaikkan akurasi dari 0.80 menjadi 0.85, spesifisitas dari 0.90 ke 0.976 [19].

Sandilya, dkk menggunakan XGBoost untuk menilai feature importance (12 variabel), lalu membandingkan algoritma XGBoost, DT, SVM, RF, LR, KNN, NB untuk melihat hasil klasifikasinya. Logistic Regression menghasilkan performansi paling baik dengan akurasi 0.87 F1-score 0.78, precision 0.82, recall 0.74 [20].

Anshori, dkk menerapkan Information Gain untuk memilih lima fitur terbaik, kemudian membandingkan hasil prediksi logistic regression dengan seluruh fitur dibandingkan dengan hasil prediksi logistic regression dan 5 fitur terbaik dari information gain. Hasilnya information gain menunjukkan adanya peningkatan akurasi dari 0.838 ke 0.8556, F-score dari 0.719 ke 0.772, dan ROC AUC dari 0.822 ke 0.869 [10].

Man berfokus pada penggunaan Logistic Regression untuk memprediksi resiko kematian pada penyakit gagal jantung hasil yang diperoleh menunjukkan bahwa model Logistic Regression yang dioptimasi dengan metode *backward optimization* menunjukkan performansi yang cukup baik dimana didapatkan nilai ROC sebesar 0.845 [21].

Jumardin, dkk membuat aplikasi prediksi kematian gagal jantung dengan sebelumnya membandingkan algoritma Random Forest, Logistic Regression, dan KNN. Hasilnya Logistic Regression memberikan performansi paling baik dengan akurasi 0.79 dan F1-score 0.7931 [22].

Arista, dkk menguji Decision Tree dan Random Forest untuk memprediksi resiko kematian pada penderita gagal jantung. Hasilnya Random Forest unggul pada akurasi 0.77, tetapi Decision Tree lebih baik dalam mendeteksi kematian dengan nilai precision 0.76 dan recall 0.52 [23].

Khan, dkk menggunakan Random Forest, XGBoost, Gradient Boost, serta ensemble (voting, stacking, blending) learning untuk melakukan prediksi resiko kematian pada pasien gagal jantung. Hasilnya Gradient Boost memberikan nilai performansi terbaik dengan akurasi sebesar 0.91. Selanjutnya Khan memakai SHAP analysis untuk melihat fitur apa saja yang paling berpengaruh pada proses prediksi [24].

### 2.2. Mutual Information sebagai Metode Seleksi Fitur

Biswas, dkk mencoba menguji 3 metode seleksi fitur yakni chi square, ANOVA, dan Mutual Information untuk memprediksi penyakit jantung pada tahap awal. Hasilnya menunjukkan bahwa Mutual Information hasil paling optimal dimana didapatkan akurasi mencapai 0.945 [25].

Ahmad, dkk menguji tiga metode seleksi fitur—Mutual Information, ANOVA, dan Chi-square—pada dataset penyakit jantung klinis dengan 11 model ML/DL. MI menunjukkan performa tertinggi, menghasilkan akurasi hingga 0.82 dan recall 0,94 pada model neural network. Bahkan Logistic Regression dan Random Forest meningkat performanya dengan MI (masing-masing sekitar 0.82 dan 0.80), sementara model sederhana seperti Naive Bayes dan Decision Tree tetap kompetitif namun kurang optimal [13].

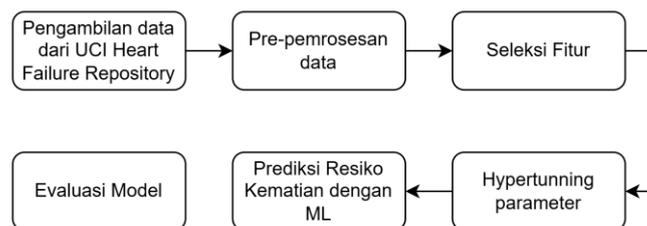
Dalam studi prediksi penyakit jantung koroner, Rehman dkk, memilih fitur terbaik berdasarkan nilai MI, kemudian melatih berbagai model (LR, RF, KNN, SVM, NB). Hasilnya, fitur yang dipilih dengan MI memberikan model yang kompetitif dengan akurasi tinggi dan AUC optimal, mendemonstrasikan sifat model-agnostik MI dan efisiensi komputasi dalam proses filter awal [26].

Xue, dkk mengembangkan varian Mutual Information dengan skema Logistic Redundancy Coefficient Gradual Upweighting (LRGU-MIFS) untuk mengatasi redundansi fitur pada sistem Internet of Medical Things (IoMT). Hasilnya, LRGU-MIFS berhasil meningkatkan F1-score deteksi intrusi menjadi 0.925 dibanding MIFS konvensional yang hanya mencapai 0.88 [1].

Berdasarkan kajian di atas Mutual Information bisa menjadi alternatif untuk menjadi salah satu metode seleksi fitur pada bidang Kesehatan. Mutual Information (MI) dipilih karena mampu mengukur ketergantungan non-linier penuh antara setiap fitur dan label dalam satu langkah perhitungan, bersifat model-agnostik, efisien secara komputasi (sebagai metode filter murni), serta terbukti stabil dalam memilih fitur meski data di-subsample. Dengan demikian, MI diharapkan dapat berfungsi sebagai tahap awal yang kuat untuk memangkas fitur atau sebagai pelengkap filter lain, guna meningkatkan akurasi, stabilitas, dan generalisasi model prediksi kematian gagal jantung.

### 3. Metode

Pada penelitian ini dataset yang digunakan diperoleh dari UCI Machine Learning Repository yang berisi 299 rekam medis pasien gagal jantung yang dikumpulkan di Faisalabad Institute of Cardiology dan Allied Hospital di Faisalabad (Punjab, Pakistan), selama April–December 2015. Secara umum penelitian yang dilakukan dimulai dari tahap pengumpulan data dari dataset, dilanjutkan dengan pre-proses data untuk memastikan data yang diperoleh dapat digunakan pada proses penelitian. Selanjutnya dikarenakan jumlah data “DEATH “ dan “LIVE” tidak seimbang maka dilakukan proses SMOTE dan bootstrapping. Setelah itu data dibagi menjadi data latih dan data uji dengan persentase 70% data latih dan 30% data uji. Kemudian dilakukan proses seleksi fitur dan hypertuning parameter untuk algoritma yang digunakan pada data latih. Hasil dari seleksi fitur dan hypertuning parameter dari data latih selanjutnya diterapkan pada data uji untuk dilakukan proses pengujian. Hasil pengujian kemudian dianalisis berdasarkan nilai akurasi, f1-score, spesifitas, dan sensitifitas. Alur kerja dari penelitian ini ditunjukkan pada Gbr.1.



Gambar 1. Alur Kerja Penelitian

#### 3.1. Data Set

Dalam penelitian ini, digunakan data dari UCI Heart Failure Clinical Records Repository, salah satu kumpulan data medis *open access* yang paling banyak digunakan untuk analisis gagal jantung [27]. Dataset ini berisi 299 catatan pasien yang didiagnosis dengan gagal jantung kelas III atau IV menurut kriteria New York Heart Association (NYHA). Setiap rekaman memuat 12 fitur klinis yang mencakup variabel demografis (usia, jenis kelamin), kondisi komorbid (anemia, hipertensi, diabetes, kebiasaan merokok), hasil laboratorium (kreatinin fosfokinase, serum kreatinin, serum natrium, jumlah trombosit), parameter jantung (fraksi ejeksi), durasi tindak lanjut (waktu follow-up dalam hari), serta label target “death\_event” yang menunjukkan apakah pasien meninggal selama periode pengamatan. Dari 299 pasien yang tercatat, 96 di antaranya meninggal dunia (32,1 %), sedangkan 203 pasien lainnya bertahan hidup (67,9 %). Data bersifat numerik kontinyu untuk nilai laboratorium dan fraksi ejeksi, serta biner untuk kondisi komorbid dan outcome. Tabel 1 menunjukkan penjelasan detail mengenai dataset yang digunakan.

**Tabel 1.** Dataset UCI Heart Failure Clinical Records Repository

No.	Fitur	Deskripsi	Rentang Nilai	Satuan/Metrik
1	Age	Usia pasien	40–95	Tahun
2	Anemia	Penurunan jumlah sel darah merah atau hemoglobin	0, 1	Boolean
3	High blood pressure	Apakah pasien menderita tekanan darah tinggi	0, 1	Boolean
4	Creatinine phosphokinase	Kadar enzim kreatinin fosfokinase dalam darah	23–7,861	Mcg/L
5	Diabetes	Apakah pasien menderita diabetes	0, 1	Boolean
6	Ejection fraction	Persentase darah yang keluar dari jantung di setiap kontraksi	14–80	Persentase (%)
7	Sex	Jenis kelamin pasien (pria atau wanita)	0, 1	Biner
8	Platelets	Jumlah trombosit dalam darah	25.01–850	Kiloplatelets/mL
9	Serum creatinine	Kadar kreatinin dalam darah	0.50–9.40	mg/dL
10	Serum sodium	Kadar natrium dalam darah	114–148	mEq/L
11	Smoking	Apakah pasien adalah perokok	0, 1	Boolean
12	Time	Lama masa tindak lanjut (follow-up)	4–285	Hari
13	Death event (target)	Apakah pasien meninggal selama masa tindak lanjut	0, 1	Boolean

### 3.2. Pra-pemrosesan Data

Sebelum membangun model klasifikasi, dilakukan tahapan pra-pemrosesan yang sistematis untuk menjamin kualitas data dan memitigasi potensi bias. Pertama, seluruh atribut diperiksa keberadaan nilai null atau tidak konsisten; meski UCI menyajikan data yang lengkap, verifikasi ini penting untuk memastikan integritas dataset dan menegaskan tidak ada nilai hilang. Kedua, setiap fitur numerik distandarisasi menggunakan transformasi z-score, sehingga data memiliki rata-rata 0 dan simpangan baku 1, memudahkan konvergensi algoritma dan membuat bobot fitur menjadi sebanding. Ketiga, untuk menangani ketidakseimbangan kelas antara pasien yang selamat dan meninggal, kami menerapkan SMOTE (Synthetic Minority Over-sampling Technique) yang menciptakan sampel sintesis pada kelas minoritas berdasarkan interpolasi tetangga terdekat, sehingga distribusi label menjadi seimbang. Keempat, guna mengukur ketahanan model terhadap variasi data sekaligus mendapatkan estimasi varians performa, dilakukan bootstrapping dengan pengambilan ulang (sampling with replacement) pada data latih untuk membentuk puluhan hingga ratusan subsample.

### 3.3. Seleksi Fitur

Dalam upaya mereduksi dimensi dan memilih atribut yang paling informatif untuk klasifikasi gagal jantung, Mutual Information (MI) digunakan sebagai metode seleksi fitur tipe filter. Secara konseptual, MI mengukur seberapa banyak informasi bersama antara fitur  $X_i$  dan label  $Y$ , dinyatakan oleh persamaan:

$$I(X_i; Y) = \sum_{x \in X_i} \sum_{y \in Y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \quad (1)$$

$P(x, y)$  = distribusi bersama

$P(x)P(y)$  = marginals

Nilai MI yang tinggi menunjukkan ketergantungan kuat—baik linier maupun non-linier—antara fitur dan outcome (mendekati 1). Implementasi praktis dilakukan dengan sebelumnya melakukan pembagian data yang sudah melewati pra-pemrosesan menjadi 70% data latih dan 30% data uji. MI kemudian diaplikasikan pada data latih menggunakan fungsi `mutual_info_classif` dari `scikit-learn`, yang secara otomatis memperkirakan MI diskrit maupun kontinu. Setelah didapatkan  $I(X_i; Y)$  dari semua fitur, dipilih 5 fitur teratas untuk selanjutnya diterapkan juga pada data uji.

### 3.4. Machine Learning

Pada penelitian ini digunakan empat algoritma klasifikasi: Random Forest, Gradient Boosting, XGBoost, dan Logistic Regression. Masing-masing algoritma memiliki karakteristik tersendiri. Untuk setiap parameter pada masing-masing algoritma dilakukan hyperparameter tuning menggunakan bayes search dengan jumlah iterasi 50 dan 5-fold cross-validation. Metrik evaluasi meliputi akurasi, precision, recall, dan F1-score. Proses tuning terus diulang hingga diperoleh konfigurasi optimal yang memaksimalkan kombinasi metrik.

#### 3.4.1. Random forest

Random Forest adalah metode ensemble yang menggabungkan banyak pohon keputusan (decision trees) untuk meningkatkan akurasi dan mengurangi risiko overfitting. Setiap pohon dibangun dari sampel bootstrap (sampling with replacement) yang berbeda dari data pelatihan, dan di setiap simpul pohon hanya dipilih subset acak dari fitur untuk mencari pemisahan terbaik. Dengan cara ini, setiap pohon “melihat” data dan fitur yang sedikit berbeda, sehingga kesalahan yang bersifat spesifik terhadap satu pohon dapat dikompensasikan oleh pohon lain dalam hutan. Prediksi akhir dibuat dengan mekanisme majority voting (untuk klasifikasi) atau rata-rata (untuk regresi) dari prediksi individual setiap pohon.

Dalam proses hypertuning Random Forest, dilakukan tuning pada lima parameter utama untuk menemukan konfigurasi terbaik. Pertama, `n_estimators` diatur dalam rentang 100 hingga 1.000 pohon, sehingga dapat mengeksplorasi pengaruh jumlah estimator terhadap stabilitas dan akurasi model. Kedua, `max_depth` dipatok antara 3 hingga 20 untuk membatasi kedalaman pohon dan mencegah overfitting, sekaligus melihat sejauh mana pemangkasan dapat mempertahankan performa. Selanjutnya, `min_samples_split` divariasikan dari 2 hingga 20, mengontrol jumlah minimum sampel yang diperlukan untuk memecah simpul, demi menyeimbangkan kompleksitas pohon. Kemudian, `min_samples_leaf` ditetapkan antara 1 hingga 10 untuk menentukan jumlah minimal observasi di setiap daun pohon, yang juga membantu meredam overfitting pada subset data kecil. Terakhir, parameter bootstrap dieksplorasi dalam dua opsi True atau False untuk membandingkan hasil model yang menggunakan sampling dengan pengembalian terhadap yang tak menggunakan pengembalian.

#### 3.4.2. Gradient Boosting

Gradient Boosting membangun model secara berurutan dimana setiap estimator baru (biasanya pohon keputusan pendek atau stump) dilatih untuk mempelajari residual (error) model sebelumnya. Ide dasarnya adalah meminimalkan fungsi loss  $L(y, F(x))$  dengan mengambil langkah menuruni gradien loss di ruang fungsi. Gradient Boosting unggul dalam menangkap pola non-linier kompleks dan interaksi fitur, tetapi rentan terhadap overfitting jika terlalu banyak iterasi atau *learning rate* terlalu besar.

Dalam hypertuning Gradient Boosting, dilakukan tuning pada enam hyperparameter utama: `n_estimators` (100–1000) untuk menentukan jumlah pohon boosting; `learning_rate`

(0,01–0,3, log-uniform) untuk mengatur kontribusi tiap pohon; max\_depth (3–10) membatasi kedalaman pohon; min\_samples\_split (2–20) dan min\_samples\_leaf (1–10) mengontrol jumlah sampel minimum saat membagi simpul dan membentuk daun; serta subsample (0,5–1,0) untuk memperkenalkan variasi data pada tiap iterasi.

**3.4.3. XGBoost**

XGBoost (eXtreme Gradient Boosting) adalah peningkatan praktis dari algoritme Gradient Boosting yang dioptimalkan untuk efisiensi dan performa. Selain prinsip dasar boosting, XGBoost menambahkan penalti L1 dan L2 pada kompleksitas model, yang membantu mencegah overfitting. XGBoost juga menerapkan *sharding* data, *column subsampling*, dan *cache-aware block structure* untuk mempercepat perhitungan pada dataset besar, serta algoritme pencarian split yang efisien. Hasilnya, XGBoost sering kali lebih cepat dan lebih akurat daripada implementasi Gradient Boosting standar.

Dalam hypertuning XGBoost, lima hyperparameter kunci yang detuning adalah: n\_estimators (100–1000) untuk jumlah pohon; learning\_rate (0,01–0,3, log-uniform) untuk skala setiap pohon; max\_depth (3–10) membatasi kompleksitas pohon; subsample (0,5–1,0) untuk sampling data per pohon; dan colsample\_bytree (0,5–1,0) untuk sampling fitur tiap pohon. Pencarian acak digunakan untuk menemukan kombinasi optimal yang memaksimalkan performa model.

**3.4.4. Logistic Regression**

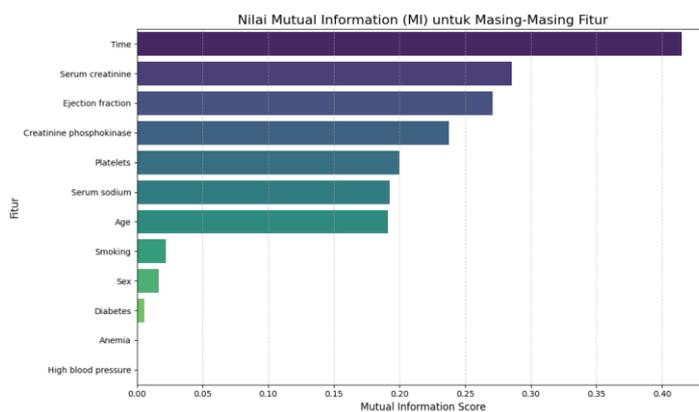
Logistic Regression adalah model statistik linier yang digunakan untuk klasifikasi biner dengan memperkirakan probabilitas kejadian suatu peristiwa. Model ini menggunakan fungsi logit untuk memetakan kombinasi linier fitur X ke probabilitas p dalam rentang [0,1]. Meskipun bersifat linier, Logistic Regression seringkali efektif ketika hubungan antara fitur dan log-odds cukup sederhana. Model ini cepat, mudah diinterpretasikan (koefisien menunjukkan arah dan kekuatan pengaruh fitur), dan dapat diperkuat dengan regularisasi L1 (LASSO) atau L2 (Ridge) untuk mencegah overfitting.

Dalam hypertuning Logistic Regression, empat hyperparameter utama yang dituning adalah: C ( $1e^{-6}$ – $1e^6$ , log-uniform) untuk mengatur kekuatan regularisasi; penalty ('l1', 'l2', 'elasticnet', 'none') untuk memilih jenis regularisasi; solver ('liblinear', 'saga') yang mendukung semua opsi penalty; serta l1\_ratio (0–1) khusus untuk penalty 'elasticnet' guna menyeimbangkan kontribusi L1 dan L2.

**4. Hasil dan Pembahasan**

Pada bagian ini dibahas mengenai hasil penelitian meliputi fitur yang terseleksi menggunakan Mutual Information, kemudian hasil dari hypertunning parameter untuk masing-masing algoritma machine learning, selanjutnya dibahas hasil prediksi resiko kematian gagal jantung menggunakan fitur hasil dari seleksi fitur dan algoritma machine learning yang menunjukkan performansi terbaik setelah dituning.

**4.1. Fitur Terseleksi Hasil Mutual Information**



**Gambar 2.** Nilai MI untuk Masing-Masing Fitur

Gbr.2 menunjukkan nilai MI untuk masing-masing fitur menggunakan data latih. Dari 12 fitur yang ada, 5 fitur dengan nilai tertinggi adalah time, serum creatinin, ejection fraction, creatinine phosphokinase, dan platelets. Lima fitur utama ini yang akan digunakan pada proses prediksi resiko kematian pada penyakit gagal jantung. Fitur yang memiliki nilai MI terbesar adalah time yakni sebesar 0.415, yang menunjukkan bahwa durasi tindak lanjut pasien (jumlah hari sejak diagnosis hingga akhir pengamatan) merupakan indikator paling informatif dalam memprediksi kemungkinan kematian akibat gagal jantung. Nilai ini mencerminkan hubungan non-linear yang kuat antara lama waktu pemantauan dan status kematian pasien. Hal ini selaras dengan prinsip dasar mutual information, yakni mengukur seberapa besar suatu fitur memberikan informasi baru terhadap label target. Dari informasi ini diketahui bahwa pasien yang bertahan lebih lama dalam periode pengamatan umumnya memiliki kondisi yang lebih stabil, sedangkan pasien dengan waktu tindak lanjut yang pendek cenderung mengalami kematian lebih awal. Dari penelitian ini menunjukkan bahwa dimensi waktu memiliki pengaruh yang cukup signifikan dalam survival analysis dan model klasifikasi klinis.

Fitur kedua yang menonjol adalah serum\_creatinine, dengan nilai mutual information sebesar 0.285. Serum creatinine mencerminkan fungsi ginjal pasien dan merupakan biomarker penting dalam manajemen gagal jantung, karena penurunan fungsi ginjal sering kali menjadi komplikasi lanjutan dari kerusakan jantung. Nilai mutual information yang tinggi menunjukkan bahwa perubahan dalam kadar kreatinin serum memiliki hubungan erat dengan tingkat risiko kematian. Secara klinis, hal ini dapat dijelaskan oleh fakta bahwa disfungsi ginjal memperburuk sirkulasi darah, meningkatkan beban kerja jantung, dan mempercepat kegagalan sistemik. Oleh karena itu, fitur ini bukan hanya relevan dari sudut pandang statistik, tetapi juga memiliki signifikansi medis yang kuat. Temuan ini sejalan dengan berbagai penelitian sebelumnya yang menyebutkan kreatinin sebagai salah satu prediktor paling kuat dalam model prediksi gagal jantung.

Selanjutnya, ejection\_fraction menduduki posisi ketiga dengan skor mutual information sebesar 0.271. Fraksi ejeksi mengukur persentase darah yang dipompa keluar dari ventrikel kiri setiap kontraksi, dan menjadi indikator langsung dari fungsi pemompaan jantung. Pasien dengan nilai ejection fraction yang sangat rendah cenderung mengalami gagal jantung kongestif yang lebih parah dan memiliki prognosis buruk. Dalam penelitian ini, tingginya nilai mutual information menandakan bahwa variasi dalam fraksi ejeksi sangat informatif dalam membedakan antara pasien yang selamat dan yang meninggal. Korelasi ini mendukung asumsi awal bahwa penurunan efisiensi pemompaan jantung menjadi faktor utama yang mempercepat kematian pada kasus gagal jantung berat.

Fitur keempat yang terpilih adalah creatinine\_phosphokinase (CPK), dengan nilai mutual information 0.237. CPK merupakan enzim yang dilepaskan ke dalam darah saat terjadi kerusakan otot, termasuk otot jantung. Peningkatan kadar CPK mencerminkan adanya kerusakan jaringan yang aktif, dan dalam konteks gagal jantung, dapat menjadi sinyal adanya infark miokard atau proses degeneratif lainnya. Walaupun secara klinis CPK bukan satu-satunya indikator utama untuk diagnosis gagal jantung, namun tingginya nilai mutual information dalam dataset ini menunjukkan bahwa CPK menyimpan informasi penting terkait outcome pasien. Hal ini mungkin terjadi karena pasien dengan gagal jantung kronis sering mengalami episode akut berulang yang dapat terdeteksi melalui lonjakan CPK. Hasil ini memungkinkan adanya pengkajian lebih dalam mengenai kemungkinan adanya relasi antara dinamika kerusakan otot sebagai bagian dari prognosis gagal jantung.

Fitur terakhir yang masuk lima besar adalah platelets, dengan skor mutual information 0.199. Jumlah trombosit dalam darah tidak secara langsung mencerminkan kinerja jantung, tetapi berkaitan dengan status koagulasi dan kemungkinan peradangan sistemik. Pada pasien gagal jantung, disfungsi endotel dan peradangan kronis dapat mengganggu produksi dan fungsi trombosit. Nilai MI yang cukup tinggi pada fitur ini menunjukkan bahwa meskipun tidak langsung, jumlah trombosit mengandung informasi signifikan dalam memprediksi risiko kematian. Hubungan ini bisa jadi mencerminkan keterlibatan sistemik yang lebih luas dalam patofisiologi gagal jantung.

## 4.2. Hasil Hypertuning Parameter Machine Learning

### 4.2.1. Random Forest

Setelah dilakukan hyperparameter tuning, diperoleh konfigurasi optimal untuk algoritma Random Forest dengan parameter `n_estimators = 100`, `max_depth = 20`,

`min\_samples\_split = 3`, `min\_samples\_leaf = 1`, dan `bootstrap = True`. Kombinasi ini menunjukkan bahwa model bekerja optimal dengan jumlah pohon yang relatif kecil (100 pohon), yang cukup untuk membangun generalisasi yang baik tanpa membutuhkan sumber daya komputasi berlebih. Kedalaman maksimum pohon yang mencapai 20 mengindikasikan bahwa pola dalam data cukup kompleks, sehingga diperlukan struktur pohon yang dalam untuk menangkap keterkaitan antar fitur. Pemisahan node hanya dilakukan jika terdapat minimal tiga sampel, dan setiap daun dapat terbentuk meskipun hanya berisi satu sampel. Ini mencerminkan bahwa model diberi keleluasaan untuk mengeksplorasi detail data dengan granularitas tinggi, namun tetap dijaga keseimbangannya melalui bootstrap sampling.

Evaluasi terhadap model Random Forest yang telah dituning menunjukkan performa yang sangat tinggi. Tabel 2 menunjukkan hasil evaluasi model Random Forest yang sudah dituning pada data latih. Akurasi keseluruhan mencapai 0.99, dan metrik F1-score untuk kedua kelas (0: bertahan hidup, 1: meninggal) masing-masing berada di angka 0.99. Precision dan recall yang hampir identik pada kedua kelas masing-masing di angka 0.98 dan 0.99, mengindikasikan bahwa model sangat akurat dalam membedakan pasien yang selamat dan yang meninggal. Ini sangat penting dalam konteks medis, terutama karena konsekuensi dari kesalahan klasifikasi dapat berdampak besar terhadap pengambilan keputusan klinis. Dengan recall yang sangat tinggi (0.99) pada kelas pasien meninggal, model terbukti mampu mengenali hampir semua kasus fatal secara benar. Hal ini memperkuat nilai model sebagai alat bantu diagnosis untuk mendeteksi pasien berisiko tinggi secara lebih dini dan tepat.

Dengan hasil ini, model Random Forest tidak hanya memberikan akurasi tinggi dalam prediksi resiko kematian gagal jantung, tetapi juga menunjukkan potensi untuk diimplementasikan dalam sistem pendukung keputusan medis (clinical decision support system).

**Tabel 2.** Hasil Klasifikasi Random Forest yang Sudah Dituning pada Data Latih

Kelas (Label)	Presisi	Recall	F1-Score	Jumlah Data
0 (Tidak Meninggal)	0.99	0.99	0.99	135
1 (Meninggal)	0.98	0.99	0.99	109
Akurasi = 0.99				

#### 4.2.2. Gradient Boost

Hasil hyperparameter tuning menunjukkan bahwa konfigurasi optimal untuk Gradient Boosting diperoleh pada `n\_estimators = 269`, `learning\_rate = 0.0454`, `max\_depth = 9`, `min\_samples\_split = 10`, `min\_samples\_leaf = 2`, dan `subsample = 0.7268`. Nilai `learning\_rate` yang relatif kecil mencerminkan strategi pembelajaran lambat namun stabil, yang umumnya mampu meningkatkan generalisasi model. Penggunaan `n\_estimators` sebanyak 269 pohon seimbang dengan nilai learning rate rendah ini, karena lebih banyak iterasi diperlukan untuk menyerap pola dalam data. Kedalaman maksimal pohon sebesar 9 menunjukkan bahwa model menangani hubungan fitur yang cukup kompleks, tetapi tetap dibatasi untuk menghindari overfitting. Parameter `min\_samples\_split = 10` dan `min\_samples\_leaf = 2` mengindikasikan bahwa model menghindari pemisahan terlalu dini dalam struktur pohon, sehingga pohon tetap cukup besar namun tidak terlalu spesifik terhadap data pelatihan.

Tabel 3 menunjukkan hasil evaluasi model Gradient Boost yang sudah dituning pada data latih. Dari sisi kinerja, model Gradient Boosting menunjukkan performa yang sangat baik dengan akurasi 0.97 dan F1-score rata-rata 0.97 pada kedua kelas. Untuk kelas 0 (pasien selamat), model mencapai precision 0.97 dan recall 0.98, sedangkan pada kelas 1 (pasien meninggal), precision dan recall masing-masing sebesar 0.97 dan 0.96. Ini menandakan bahwa model mampu melakukan klasifikasi secara seimbang tanpa terlalu bias terhadap salah satu kelas. Hasil macro dan weighted average yang konsisten pada angka 0.97 memperkuat bahwa performa ini stabil di seluruh distribusi kelas. Model ini cukup andal dalam mengenali pasien yang berisiko tinggi, tanpa mengorbankan ketepatan dalam klasifikasi pasien yang selamat.

**Tabel 3.** Hasil Klasifikasi Gradient Boost yang Sudah Dituning pada Data Latih

Kelas (Label)	Presisi	Recall	F1-Score	Jumlah Data
0 (Tidak Meninggal)	0.97	0.98	0.97	135
1 (Meninggal)	0.97	0.96	0.97	109
Akurasi = 0.97				

### 4.2.3. XG Boost

Hasil hyperparameter tuning untuk XGBoost menunjukkan bahwa kombinasi optimal diperoleh pada parameter `n_estimators = 452`, `learning_rate = 0.0331`, `max_depth = 7`, `subsample = 0.9975`, dan `colsample_bytree = 0.5708`. Nilai `n_estimators` yang relatif besar (452 pohon) menandakan bahwa model memerlukan jumlah pohon yang cukup banyak untuk mencapai konvergensi, yang umum terjadi saat `learning_rate` disetel rendah. Dalam hal ini, `learning_rate = 0.0331` mengindikasikan bahwa proses pembelajaran dilakukan secara perlahan, sehingga setiap pohon hanya memberi kontribusi kecil. Pendekatan ini dikenal dapat meningkatkan generalisasi model karena menghindari pembelajaran yang terlalu agresif terhadap data latih. Kedalaman pohon (`max_depth = 7`) menunjukkan bahwa model cukup dalam untuk menangkap interaksi antar fitur yang kompleks, tetapi tidak terlalu dalam hingga menyebabkan overfitting.

Parameter `subsample = 0.9975` menunjukkan bahwa hampir seluruh data digunakan dalam pelatihan setiap pohon, yang memperkuat kestabilan model meskipun mengurangi efek regularisasi dari subsampling. Namun, regularisasi tetap dikompensasi oleh nilai `colsample_bytree = 0.5708`, yang berarti hanya sekitar 57% fitur yang dipertimbangkan untuk setiap pohon. Teknik ini memaksa XGBoost untuk tidak bergantung pada fitur tertentu, sehingga meningkatkan keragaman antar pohon dalam ensemble dan mencegah overfitting. Kombinasi dari learning rate rendah dan strategi subsampling sebagian fitur membuat model belajar secara hati-hati dan robust terhadap fluktuasi data.

Tabel 4 menunjukkan hasil evaluasi model XGBoost yang sudah dituning pada data latih. Akurasi model mencapai 0.98, dengan nilai F1-score keseluruhan 0.9691, menunjukkan keseimbangan yang sangat baik antara presisi dan recall. Untuk kelas 0 (pasien bertahan), precision adalah 0.97 dan recall mencapai 0.99, sedangkan untuk kelas 1 (pasien meninggal), precision sebesar 0.98 dan recall 0.96. Meskipun ada sedikit penurunan recall pada kelas 1 dibanding Random Forest, nilai precision yang lebih tinggi menunjukkan bahwa model cenderung lebih selektif dalam memprediksi kematian, dengan lebih sedikit false positives. Macro average dan weighted average pada semua metrik evaluasi juga stabil di angka 0.98, menandakan konsistensi performa antar kelas.

**Tabel 4.** Hasil Klasifikasi XGBoost yang Sudah Dituning pada Data Latih

Kelas (Label)	Presisi	Recall	F1-Score	Jumlah Data
0 (Tidak Meninggal)	0.97	0.99	0.98	135
1 (Meninggal)	0.98	0.96	0.97	109
Akurasi = 0.98				

### 4.2.4. Logistic Regression

Hasil tuning menunjukkan bahwa konfigurasi terbaik untuk model Logistic Regression diperoleh dengan parameter `C = 1.265`, `penalty = 'l1'`, `solver = 'saga'`, dan `l1_ratio = 0.0`. Parameter `C` mengontrol kekuatan regularisasi, di mana nilai yang lebih tinggi berarti regularisasi lebih lemah. Nilai C yang berada sedikit di atas 1 menunjukkan bahwa model memberikan bobot cukup besar pada data pelatihan dan tidak terlalu membatasi kompleksitas model, meskipun masih dalam batas regularisasi yang stabil. Penggunaan regularisasi L1 (Lasso) bertujuan untuk mendorong sparsity dalam model, yaitu hanya mempertahankan fitur-fitur yang paling berpengaruh, sedangkan fitur lain akan dikurangi kontribusinya hingga nol. Solver `saga` dipilih karena mendukung penalti L1 dan bekerja baik pada dataset besar dan fitur yang tidak terstandarisasi sempurna. Meskipun `l1_ratio` hanya berlaku untuk regularisasi elasticnet, nilainya yang 0.0 mengonfirmasi bahwa model sepenuhnya menggunakan penalti L1.

Tabel 5 menunjukkan hasil evaluasi model Logistic Regression yang sudah dituning pada data latih. Dari sisi performa, model Logistic Regression mencatat akurasi sebesar 0.80 dan F1-score keseluruhan 0.788, yang berarti secara umum model masih cukup baik dalam mengenali pola antara fitur dan label pada dataset gagal jantung. Precision dan recall masing-masing berada di angka moderat: precision 0.85 dan recall 0.78 untuk kelas 0 (bertahan hidup), serta precision 0.75 dan recall 0.83 untuk kelas 1 (kematian). Meskipun tidak seimbang, model ini mampu menangkap pasien berisiko cukup baik (recall 0.83), namun dengan harga menurunnya precision, yang berarti terdapat false positives yang lebih banyak. Nilai macro dan weighted average yang sama (0.80) menunjukkan performa model stabil di dua kelas,

namun masih tertinggal cukup jauh dibandingkan model ensemble seperti Random Forest, XGBoost, atau Gradient Boosting yang memiliki F1-score di atas 0.95.

Hasil ini mencerminkan keterbatasan model Logistic Regression dalam menangani hubungan non-linier kompleks yang umum terjadi dalam data medis. Meskipun algoritma ini sederhana, cepat, dan mudah diinterpretasikan, Logistic Regression bergantung penuh pada asumsi linieritas antara kombinasi fitur dan logit dari probabilitas kelas. Oleh karena itu, ketika fitur saling berinteraksi atau memiliki pola yang tidak linier, performa Logistic Regression cenderung menurun. Namun demikian, penggunaan regularisasi L1 telah membantu model ini menghindari overfitting dan mempertahankan fitur penting secara otomatis. Model ini masih berguna sebagai baseline awal atau komponen dalam algoritma ensemble, tetapi untuk klasifikasi klinis dengan akurasi tinggi, hasil ini menunjukkan metode lain lebih unggul dalam konteks prediksi gagal jantung berbasis data klinis.

**Tabel 5.** Hasil Klasifikasi Logistic Regression yang Sudah Dituning pada Data Latih

Kelas (Label)	Presisi	Recall	F1-Score	Jumlah Data
0 (Tidak Meninggal)	0.85	0.78	0.81	135
1 (Meninggal)	0.75	0.83	0.79	109
Akurasi = 0.80				

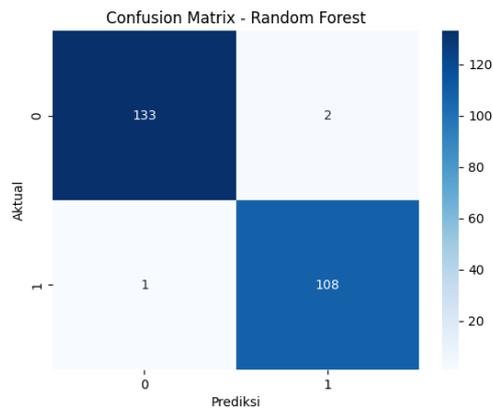
### 4.3. Hasil Prediksi Kematian Random Forest pada Data Uji

Selanjutnya dilakukan pelatihan dan pengujian menggunakan algoritma Random Forest dengan parameter hasil hypertuning. Hasil prediksi resiko kematian dari Random Forest pada Tabel 6 menunjukkan performa luar biasa berdasarkan data uji yang digunakan. Dengan akurasi sebesar 0.99, model hampir sempurna dalam memisahkan pasien yang bertahan hidup (label 0) dan yang mengalami kematian (label 1). Laporan klasifikasi menunjukkan nilai precision, recall, dan F1-score yang sangat tinggi dan seimbang untuk kedua kelas: semua berada di angka 0.98–0.99. Khususnya untuk kelas 1 (kematian), nilai recall 0.99 sangat penting karena menunjukkan bahwa hampir semua kasus kematian berhasil diidentifikasi dengan benar oleh model. Ini sangat mendukung hipotesis awal penelitian, yaitu bahwa dengan seleksi fitur yang tepat (menggunakan Mutual Information) dan tuning parameter model yang optimal, performa klasifikasi gagal jantung dapat ditingkatkan secara signifikan.

**Tabel 6.** Hasil Prediksi Kematian Menggunakan Random Forest yang Sudah Dituning pada Data Uji

Kelas	Presisi	Recall	F1-Score	Jumlah Sampel
0 (Selamat)	0.99	0.99	0.99	135
1 (Meninggal)	0.98	0.99	0.99	109
Akurasi = 0.99				

Confusion matrix pada Gbr.3 mendukung metrik tersebut secara visual. Dari total 244 data hanya tiga kesalahan klasifikasi terjadi: dua pasien yang selamat diklasifikasikan salah sebagai meninggal (false positive) dan satu pasien yang meninggal diklasifikasikan sebagai selamat (false negative). Kesalahan minimum ini sangat signifikan dalam konteks klinis karena menurunkan potensi misdiagnosis. Lebih penting lagi false negative yang hanya satu kasus menunjukkan bahwa model ini sangat efektif dalam mendeteksi pasien berisiko tinggi yang secara medis merupakan prioritas utama. Model yang dapat secara akurat memprediksi kematian akibat gagal jantung bermanfaat untuk intervensi lebih awal dan pengambilan keputusan berbasis risiko.



**Gambar 3.** Confusion Matrix Hasil Prediksi Kematian Random Forest (dengan parameter yang Sudah Dituning)

Dari sisi konfigurasi parameter `max_depth=20` memberi model kemampuan untuk membangun pohon yang cukup dalam dan mampu menangkap pola-pola kompleks dalam data klinis. Pengaturan `min_samples_split=3` dan `min_samples_leaf=1` memberi fleksibilitas pada pohon untuk membagi data bahkan dalam jumlah kecil yang mendukung identifikasi pola minoritas misalnya kasus-kasus fatal yang jumlahnya lebih sedikit. Pemilihan `n_estimators=100` terbukti cukup optimal untuk mencapai kestabilan prediksi tanpa memperbesar waktu pelatihan secara signifikan. Sementara itu penggunaan `bootstrap=True` memastikan bahwa setiap pohon dilatih dengan subset data yang berbeda mendorong variasi antar pohon dan meningkatkan kemampuan generalisasi model.

Hasil ini juga mengonfirmasi kekuatan pendekatan ensemble seperti Random Forest dibanding model linier seperti Logistic Regression. Sementara Logistic Regression mencatat F1-score hanya sekitar 0.788 (hasil hypertuning parameter) Random Forest mencapai skor 0.99 di kedua kelas. Perbedaan ini dapat dijelaskan oleh fakta bahwa Random Forest mampu menangani hubungan non-linier antar fitur dan interaksi kompleks antar variabel medis yang tidak bisa dijelaskan dengan model linier sederhana. Selain itu Random Forest secara inheren menyediakan fitur penting (feature importance) yang dapat dipakai untuk interpretasi lebih lanjut dan validasi klinis terhadap fitur-fitur utama yang memengaruhi klasifikasi.

Hasil ini juga mengonfirmasi hipotesis awal dalam penelitian bahwa seleksi fitur berbasis Mutual Information mampu menyaring atribut yang paling relevan sehingga memperkuat struktur model dan mengurangi noise. Kombinasi antara pemilihan fitur yang efisien dan penyetelan parameter yang optimal menghasilkan model klasifikasi yang tidak hanya akurat secara numerik tetapi juga seimbang dari sisi sensitivitas dan presisi. Hal ini terbukti dari tingginya nilai macro average dan weighted average pada semua metrik evaluasi. Model tidak bias terhadap salah satu kelas dan performanya stabil di seluruh distribusi data.

Secara implisit performa tinggi ini juga mencerminkan kualitas data yang telah diproses dengan baik sebelumnya. Penerapan SMOTE pada tahap pra-pemrosesan membantu mengatasi ketimpangan distribusi kelas sementara proses standarisasi dan bootstrapping memperkuat kestabilan model. Keberhasilan tuning ini menunjukkan bahwa meskipun dataset Heart Failure dari UCI relatif kecil pendekatan yang sistematis dan berbasis teori dalam pemrosesan seleksi fitur serta tuning model dapat menghasilkan prediktor yang sangat kompeten. Model ini sangat potensial untuk diterapkan lebih lanjut sebagai bagian dari sistem pendukung keputusan di rumah sakit atau aplikasi prediktif berbasis data terutama dalam membantu dokter mengidentifikasi pasien berisiko tinggi dan memprioritaskan intervensi secara lebih efektif.

### 5. Perbandingan

Table 7 menunjukkan hasil penelitian prediksi resiko kematian pasien gagal jantung beberapa tahun terakhir dan penelitian. Dari Tabel 7, penelitian ini menunjukkan performa prediktif yang sangat tinggi dengan akurasi mencapai 0.99 dan F1-score sebesar 0.99 yang secara umum lebih tinggi dibandingkan hasil dari penelitian-penelitian sebelumnya. Sebagai contoh studi oleh Abid Ishaq dan Saima Sadiq mencatat akurasi sebesar 0.926 menggunakan teknik balancing SMOTE serta berbagai algoritma seperti Extra Trees Classifier Random Forest dan AdaBoost. Meskipun pendekatan mereka telah mencakup strategi balancing dan seleksi fitur

berbasis random forest metode seleksi fitur yang mereka gunakan belum secara spesifik menangkap hubungan non-linier kompleks seperti yang dilakukan dalam penelitian ini melalui penerapan Mutual Information. Dengan demikian Penggunaan Mutual Information sebagai metode seleksi fitur untuk meningkatkan optimasi hasil prediksi risiko kematian penderita gagal jantung menjadikannya state-of-the-art dari penelitian ini.

Selanjutnya bila dibandingkan dengan pendekatan oleh Anshori, dkk yang menggunakan Information Gain penelitian ini memperlihatkan hasil yang lebih konsisten dan tinggi. Anshori berhasil meningkatkan akurasi Logistic Regression dari 0.838 menjadi 0.856 dengan F1-score meningkat ke 0.772. Meskipun Information Gain efektif dalam memfilter fitur metode ini hanya mengukur korelasi linier antara fitur dan target. Di sisi lain Mutual Information mampu mengenali keterkaitan non-linier yang lebih sesuai dengan kompleksitas data klinis seperti gagal jantung. Oleh karena itu MI memberikan fleksibilitas dan presisi lebih tinggi dalam proses seleksi fitur untuk mendukung model klasifikasi medis yang akurat dan efisien.

Studi oleh Khan menggunakan teknik explainable AI melalui SHAP analysis serta pendekatan ensemble untuk meningkatkan akurasi prediksi hingga 0.91. Meskipun pendekatan tersebut memberikan interpretabilitas yang sangat baik dan bermanfaat dalam konteks klinis performa klasifikasinya belum melampaui capaian yang diperoleh dalam penelitian ini. Kombinasi Mutual Information dan algoritma seperti Random Forest XGBoost serta Gradient Boosting yang digunakan dalam penelitian ini berhasil mengoptimalkan baik sisi akurasi maupun kestabilan antar kelas. Ini memperkuat posisi penelitian ini dalam memberikan kontribusi nyata pada pengembangan model prediksi risiko medis yang seimbang antara keandalan teknis dan efisiensi fitur.

Selain itu pendekatan minimalis yang diadopsi oleh Chicco yang hanya menggunakan dua fitur (serum creatinine dan ejection fraction) menghasilkan akurasi 0.585 dengan F1-score 0.754. Chicco menunjukkan bahwa fitur tertentu memiliki pengaruh besar terhadap prediksi kelangsungan hidup namun pendekatannya belum sepenuhnya mengoptimalkan informasi yang tersedia dalam dataset. Dalam penelitian ini pemilihan lima fitur terbaik melalui Mutual Information membuktikan bahwa model dapat tetap ringan namun sangat akurat karena fitur-fitur yang digunakan telah terbukti memiliki kontribusi informasi yang signifikan terhadap variabel target.

Dari sisi penanganan data tidak seimbang beberapa studi sebelumnya telah menerapkan SMOTE [6] atau Balanced Random Forest [7] untuk meningkatkan performa model yang menghasilkan akurasi berkisar 0.76-0.79. Penelitian ini mengembangkan pendekatan tersebut dengan mengombinasikan SMOTE dan bootstrapping untuk penyeimbangan kelas serta Mutual Information untuk pemilihan fitur yang secara bersama-sama berkontribusi terhadap peningkatan akurasi dan kestabilan prediksi. Dengan demikian kombinasi metode balancing dan seleksi fitur yang dirancang terbukti berperan penting dalam membangun sistem klasifikasi medis yang lebih tangguh.

Akhirnya dari segi jumlah fitur yang digunakan penelitian ini menyeimbangkan antara kompleksitas model dan performa prediktif. Hanya dengan menggunakan lima fitur hasil seleksi Mutual Information model telah mampu mencapai kinerja yang sangat tinggi. Sebagai perbandingan penelitian oleh Khan menggunakan 12 fitur dan menghasilkan akurasi 0.91 namun dengan kompleksitas model yang lebih tinggi. Hal ini menunjukkan bahwa peningkatan jumlah fitur tidak selalu menjamin peningkatan performa prediksi secara signifikan. Oleh karena itu pendekatan yang diterapkan dalam penelitian ini tidak hanya akurat tetapi juga efisien secara komputasi dan lebih mudah diimplementasikan dalam praktik klinis. Dengan keseluruhan strategi yang dirancang penelitian ini memberikan kontribusi penting dalam pengembangan sistem pendukung keputusan berbasis machine learning untuk prediksi risiko kematian pada pasien gagal jantung.

**Tabel 7.** Perbandingan dengan Penelitian Sebelumnya

No	Tahun	Metode Balancing Data	Metode Seleksi Fitur	Jumlah Fitur	Algoritma Machine Learning	Performansi
1	2020[5]		Analisis biostatistik dan random forest.	2	RF	Akurasi = 0.585 F1-score = 0.754

2	2020[6]	SMOTE+ ENN	<ul style="list-style-type: none"> <li>•SymmetricalUncertAttributeEvaluator (SU)</li> <li>•InfoGainAttributeEvaluator (IG)</li> <li>•GainRatioAttributeEvaluator (GR)</li> <li>•CorrelationAttributeEvaluator (CO)</li> </ul>	2	RF, SVM, KNN	Akurasi = 0.7947 F-measure = 0.792
3	2021[16]	SMOTE	Random forest	9	RF, DT, AdaBoost, LR, SGD, GBM, ETC, GNB, SVM	Akurasi = 0.9262
4	2021[7]	Balanced random Forest (BRF)	Chi-Square dan RFE	9	RF, SVM, KNN, LR, AdaBoost	Akurasi = 0.7625
5	2023[8]		Integrasi cox proportional hazard model dan Random Forest	3	MLP	Akurasi = 0.754
6	2024[10]		Information Gain	5	LR	Akurasi = 0.8556 F-score = 0.772
7	2025[24]	SMOTE+ bootstrapping		12	RF, XGBoost, GB, VC, SC, BC	Akurasi = 0.9139
8	2025- Penelitian ini	SMOTE+ bootstrapping	Mutual Information	5	RF, XGBoost, GB, LR	Akurasi = 0.987 F1-score = 0.99

### 6. Kesimpulan

Penelitian ini telah berhasil mengembangkan model prediksi risiko kematian pada penderita gagal jantung menggunakan pendekatan seleksi fitur berbasis Mutual Information (MI) dan beberapa algoritma machine learning yaitu Random Forest XGBoost Gradient Boosting dan Logistic Regression. Hasil eksperimen menunjukkan bahwa kombinasi Mutual Information dan Random Forest menghasilkan performa tertinggi dengan akurasi 0.99 dan F1-score 0.99 menegaskan efektivitas MI dalam menyeleksi fitur yang relevan secara statistik dan klinis. Temuan ini mendukung tujuan penelitian dalam mengoptimalkan klasifikasi dengan fitur yang lebih sedikit namun bermakna sekaligus mengatasi kompleksitas data medis yang bersifat non-linier. Implikasi dari penelitian ini sangat penting bagi pengembangan sistem pendukung keputusan klinis terutama di fasilitas layanan kesehatan dasar karena model yang dihasilkan bersifat efisien akurat dan mudah diimplementasikan. Meskipun demikian penelitian ini memiliki keterbatasan pada penggunaan satu sumber dataset (UCI Heart Failure) yang mungkin belum mewakili variasi populasi lebih luas. Oleh karena itu penelitian lanjutan disarankan untuk menguji model pada dataset klinis nyata yang lebih besar dan heterogen serta mengeksplorasi pendekatan explainable AI guna meningkatkan interpretabilitas model bagi tenaga medis.

**Kontribusi Penulis:** Konseptualisasi: Elsa Sari Hayunah Nurdiniyah; Metodologi: XX; Perangkat Lunak: Aisya Nur Aulia Yusuf; Validasi: Elsa Sari Hayunah Nurdiniyah Aisya Nur Aulia Yusuf dan Rasyida Shabihah Zukron Aini; Analisis formal: Rahardian Luthfi Prasetyo; Investigasi: Elsa Sari Hayunah Nurdiniyah; Sumber daya: Aisya Nur Aulia Yusuf; Kurasi data: Rasyida Shabihah Zukron Aini; Penulisan—persiapan draf asli: Elsa Sari Hayunah

Nurdiniyah; Penulisan—peninjauan dan penyuntingan: Aisya Nur Aulia Yusuf; Visualisasi: Rasyida Shabihah Zukron Aini; Supervisi: Rahardian Luthfi Prasetyo

**Pendanaan:** Penelitian ini tidak menerima pendanaan eksternal.

**Pernyataan Ketersediaan Data:** Penulis mengambil data dari UCI Heart Failure Clinical Records Repository yang bisa diakses pada:

<https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records>

**Ucapan Terima Kasih:** Penulis mengucapkan terima kasih kepada seluruh pihak yang membantu dalam keberhasilan penelitian ini.

**Konflik Kepentingan:** Penulis menyatakan tidak ada konflik kepentingan.

## Referensi

- [1] S. L. James *dkk.*, “Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017,” *The Lancet*, vol. 392, no. 10159, hlm. 1789–1858, Nov 2018, doi: 10.1016/s0140-6736(18)32279-7.
- [2] E. J. Benjamin *dkk.*, “Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association,” *Circulation*, vol. 139, no. 10, Mar 2019, doi: 10.1161/CIR.0000000000000659.
- [3] Tim Riskesdas 2018, “Laporan Riskesdas 2018 Nasional,” Lembaga Penerbit dan Pengembangan Kesehatan, Jakarta, 2018, 2019.
- [4] “Heart Failure - What Is Heart Failure? | NHLBI, NIH.” Diakses: 18 Juli 2025. [Daring]. Tersedia pada: <https://www.nhlbi.nih.gov/health/heart-failure>
- [5] D. Chicco dan G. Jurman, “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone,” *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, Des 2020, doi: 10.1186/s12911-020-1023-5.
- [6] O. O. Oladimeji dan O. Oladimeji, “Predicting Survival of Heart Failure Patients Using Classification Algorithms,” *JITCE J. Inf. Technol. Comput. Eng.*, vol. 4, no. 02, hlm. 90–94, Sep 2020, doi: 10.25077/jitce.4.02.90-94.2020.
- [7] A. Newaz, N. Ahmed, dan F. Shahriyar Haq, “Survival prediction of heart failure patients using machine learning techniques,” *Inform. Med. Unlocked*, vol. 26, hlm. 100772, 2021, doi: 10.1016/j.imu.2021.100772.
- [8] Z. Wu, Y. Wu, X. He, L. Wan, L.-C. Tsai, dan A. Chen, “Application of Machine Learning for Heart Failure Prediction,” dalam *2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, China: IEEE, Mei 2023, hlm. 276–282. doi: 10.1109/icaibd57115.2023.10206268.
- [9] A. M. Qadri, M. S. A. Hashmi, A. Raza, S. A. J. Zaidi, dan A. U. Rehman, “Heart failure survival prediction using novel transfer learning based probabilistic features,” *PeerJ Comput. Sci.*, vol. 10, hlm. e1894, Mar 2024, doi: 10.7717/peerj-cs.1894.
- [10] M. Anshori, M. S. Haris, dan A. Wahyudi, “Logistic Regression’s Effectiveness in Feature Selection with Information Gain in Predicting Heart Failure Patients,” *J. Enhanc. Stud. Inform. Comput. Appl.*, vol. 1, no. 2, hlm. 35–39, Jul 2024, doi: 10.47794/jesica.v1i2.8.
- [11] B. Chulde-Fernández *dkk.*, “Classification of Heart Failure Using Machine Learning: A Comparative Study,” *Life*, vol. 15, no. 3, hlm. 496, Mar 2025, doi: 10.3390/life15030496.
- [12] S. Liu dan M. Motani, “Improving Mutual Information based Feature Selection by Boosting Unique Relevance,” 17 Desember 2022, *arXiv: arXiv:2212.06143*. doi: 10.48550/arXiv.2212.06143.
- [13] B. Ahmad, J. Chen, dan H. Chen, “Feature selection strategies for optimized heart disease diagnosis using M L and DL models”, doi: 10.48550/arXiv.2503.16577.
- [14] H. Roubhi, A. Hacine Gharbi, K. Rouabah, dan P. Ravier, “Mutual Information-based Feature Selection Strategy for Speech Emotion Recognition using Machine Learning Algorithms Combined with the Voting Rules Method,” *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, hlm. 19207–19213, Feb 2025, doi: 10.48084/etasr.9066.
- [15] F. Souza, C. Premebida, dan R. Araújo, “High-Order Conditional Mutual Information Maximization for dealing with High-Order Dependencies in Feature Selection,” *Pattern Recognit.*, vol. 131, hlm. 108895, Nov 2022, doi: 10.1016/j.patcog.2022.108895.
- [16] A. Ishaq *dkk.*, “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques,” *IEEE Access*, vol. 9, hlm. 39707–39716, 2021, doi: 10.1109/access.2021.3064084.
- [17] S. M. M. Zaman, W. M. Qureshi, Md. M. S. Raihan, A. B. Shams, dan S. Sultana, “Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm,” dalam *2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, Dhaka, Bangladesh: IEEE, Des 2021, hlm. 117–120. doi: 10.1109/wiecon-ece54711.2021.9829577.
- [18] S. Adi dan A. Wintarti, “KOMPARASI METODE SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBORS (KNN), DAN RANDOM FOREST (RF) UNTUK PREDIKSI PENYAKIT GAGAL JANTUNG,” *MATHunesa J. Ilm. Mat.*, vol. 10, no. 2, hlm. 258–268, Jul 2022, doi: 10.26740/mathunesa.v10n2.p258-268.
- [19] B. F. Sitanggang dan P. Sitompul, “Deteksi Awal Kelangsungan Hidup Pasien Gagal Jantung Menggunakan MachineLearningMetode RandomForest”.
- [20] P. K. Sandilya, D. Pal, S. R. Dahiya, dan S. K. Dana, “Prognostic Modeling for Heart Failure Survival: A Classification Approach,” dalam *2024 11th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India: IEEE, Mar 2024, hlm. 415–420. doi: 10.1109/spin60856.2024.10512260.
- [21] H. Man, “Research on Predicting Survival in Heart Failure Patients of Logistic Regression Models,” *Theor. Nat. Sci.*, vol. 53, no. 1, hlm. 255–261, Nov 2024, doi: 10.54254/2753-8818/53/20240234.

- 
- [22] Jumardin dan H. Noprisson, "Perancangan Prototipe Aplikasi Prediksi Kematian Akibat Gagal Jantung Menggunakan Metode Machine Learning Berdasarkan Data Heart Failure Clinical Records," vol. 7, no. 3, 2024, doi: 10.36085.
- [23] Y. Arista, N. Sulistianingsih, T. Fitriasia, dan Z. D. Febriyanzi, "Prediksi Gagal Jantung: Pendekatan Data Mining untuk Deteksi dan Penilaian Risiko".
- [24] N. A. Khan, Md. F. Bin Hafiz, dan Md. A. Pramanik, "Enhancing predictive modelling and interpretability in heart failure prediction: a SHAP-based analysis," *Int. J. Inform. Commun. Technol. IJ-ICT*, vol. 14, no. 1, hlm. 11, Apr 2025, doi: 10.11591/ijict.v14i1.pp11-19.
- [25] N. Biswas *dkk.*, "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," *BioMed Res. Int.*, vol. 2023, no. 1, Jan 2023, doi: 10.1155/2023/6864343.
- [26] M. U. Rehman *dkk.*, "Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment," *Sci. Rep.*, vol. 15, no. 1, Apr 2025, doi: 10.1038/s41598-025-96437-1.
- [27] "Heart Failure Clinical Records." UCI Machine Learning Repository, Online, 2020. [Daring]. Tersedia pada: <https://doi.org/10.24432/C5Z89R>