Perbandingan Algoritma Machine Learning Untuk Deteksi Gagal Jantung Berbasis Seleksi Fitur RFECV dan Penyeimbangan Data SMOTE

Ari Setyawan, Neny Sulistianingsih, Ria Rismayati

Universitas Bumigora, Mataram, Indonesia

Correspondence : *e-mail*: arisetyawanari7921@gmail.com

Abstrak

Deteksi dini gagal jantung merupakan tantangan signifikan dalam dunia medis karena kompleksitas faktor risikonya. Penelitian ini bertujuan membandingkan kinerja enam algoritma machine learning dalam memprediksi risiko gagal jantung dengan pendekatan CRISP-DM. Data klinis sebanyak 299 pasien diproses melalui seleksi fitur menggunakan Recursive Feature Elimination with Cross-validation (RFECV) serta penyeimbangan kelas dengan Synthetic Minority Over-sampling Technique (SMOTE). Algoritma yang dievaluasi meliputi Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, dan Gradient Boosting. Evaluasi dilakukan menggunakan validasi silang berstrata dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan Random Forest mencapai performa terbaik dengan akurasi dan F1-score sebesar 91,20%, diikuti Gradient Boosting dengan 90,20%. Implementasi SMOTE terbukti meningkatkan kemampuan model, terutama dalam mendeteksi kelas minoritas. Temuan ini menegaskan bahwa metode ensemble seperti Random Forest, dikombinasikan dengan RFECV dan SMOTE, efektif untuk klasifikasi risiko gagal jantung secara akurat dan andal.

Kata kunci: *gagal jantung, machine learning, Random Forest, Gradient Boosting, SMOTE.*

Abstract

Early detection of heart failure remains a significant challenge in the medical field due to the complexity of its risk factors. This study aims to compare the performance of six machine learning algorithms in predicting heart failure risk using the CRISP-DM framework. A clinical dataset of 299 patients was processed through feature selection with Recursive Feature Elimination with Cross-validation (RFECV) and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE). The evaluated algorithms include Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, and Gradient Boosting. Model performance was assessed using stratified cross-validation with accuracy, precision, recall, and F1-score as evaluation metrics. Results indicate that Random Forest achieved the best performance with an accuracy and F1-score of 91.20%, followed by Gradient Boosting with 90.20%. The implementation of SMOTE significantly improved model performance, particularly in detecting the minority class. These findings confirm that ensemble methods such as Random Forest, combined with RFECV and SMOTE, provide an effective and reliable solution for accurate heart failure risk classification.

Keywords*: heart failure, machine learning, Random Forest, Gradient Boosting, SMOTE.*

1. Pendahuluan

Gagal jantung merupakan salah satu penyakit kardiovaskular dengan prevalensi global yang terus meningkat dan saat ini menjangkit lebih dari 64 juta orang di seluruh dunia [1]. Kondisi ini tidak hanya menyebabkan morbiditas tinggi tetapi juga menjadi salah satu penyebab utama kematian, dengan sekitar 50% pasien dilaporkan meninggal dalam lima tahun setelah diagnosis [2]. Deteksi dini menjadi kunci untuk meningkatkan kualitas hidup pasien, namun gejalanya sering tumpang tindih dengan penyakit lain dan dipengaruhi oleh kompleksitas faktor risiko. Seiring dengan meningkatnya ketersediaan data klinis,

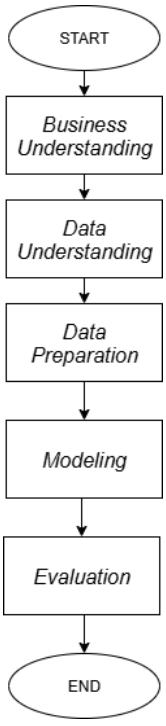
machine learning (ML) telah muncul sebagai pendekatan yang menjanjikan untuk mengidentifikasi pola-pola kompleks yang sulit dikenali metode diagnostik tradisional [3] [4].

Penelitian terdahulu menunjukkan bahwa teknik penyeimbangan data seperti Synthetic Minority Over-sampling Technique (SMOTE) dapat meningkatkan akurasi algoritma ensemble pada data yang tidak seimbang [4] [5]. Selain itu, seleksi fitur menggunakan *Recursive Feature Elimination with Cross-validation* (RFECV) terbukti mampu meningkatkan performa algoritma Gradient Boosting [6], [7]. Namun demikian, masih terdapat keterbatasan dalam penelitian yang secara sistematis mengintegrasikan seleksi fitur cerdas dengan penyeimbangan kelas dalam satu pipeline terpadu untuk evaluasi lintas berbagai algoritma.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan menganalisis dan membandingkan kinerja enam algoritma ML, yaitu Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, dan Gradient Boosting dalam mendeteksi risiko gagal jantung. Kontribusi utama dari penelitian ini adalah pembuktian bahwa pipeline pra-pemrosesan data yang terstruktur menggabungkan Recursive Feature Elimination with Cross-validation (RFECV) untuk seleksi fitur dan SMOTE untuk penyeimbangan kelas merupakan pendekatan efektif untuk memaksimalkan akurasi prediksi, khususnya pada algoritma ensemble.

2. Metode Penelitian

Penelitian ini mengadopsi kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) yang terdiri dari enam tahap, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [8] [9]. Tapi pada penelitian ini hanya menggunakan 5 tahapan saja yaitu business understanding, data understanding, data preparation, modeling dan evaluation, yang dapat dilihat pada Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Flowchart CRISP-DM

Tahap *business understanding* berfokus pada identifikasi masalah penelitian, yaitu bagaimana meningkatkan akurasi prediksi gagal jantung. Tahap data understanding dilakukan dengan mengeksplorasi dataset Heart Failure Clinical Records yang terdiri dari 299 data pasien dengan 13 atribut klinis [10] [11]. Pada tahap data preparation, dilakukan feature selection menggunakan RFECV untuk memilih atribut paling relevan, standarisasi menggunakan StandardScaler, serta penyeimbangan kelas dengan SMOTE [12]. Tahap modeling menggunakan enam algoritma machine learning, yaitu Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), dan Gradient Boosting (GB). Tahap evaluation menggunakan skema 5-Fold Stratified Cross-validation dengan metrik akurasi, presisi, recall, dan F1-score sebagai ukuran performa [8][9]. Kemudian Melakukan pengujian tambahan dengan dua dataset berbeda sebagai penguat hasil penelitian.

2.1 Decision Tree (DT)

Decision Tree membagi data berdasarkan atribut yang memaksimalkan informasi [13]. Fungsi entropi didefinisikan sebagai:

$$\mathrm{H}(S) = -\sum_{i=1}^{n} p_1 log_2 p_i$$

Dimana:
- (c) adalah jumlah kelas,
- (p_i) adalah probabilitas kemunculan kelas ke-i pada himpunan data S.

Atribut terbaik dipilih menggunakan Information Gain:

$$\mathrm{IG}(S, A) = \mathrm{H}(S) - \sum_{v \in \mathrm{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

Dimana:
- (S) Adalah himpunan data
- (A) adalah atribut yang diuji
-$\big(Values(A)\big)$ Adalah himpunan nilai yang mungkin dari atribut A
$-$ (S_v) adalah subset data S untuk nilai atribut A sama dengan v
- (|S|) adalah jumlah total data dalam himpunan S
- ($|S_v|$) adalah jumlah data dalam subset S_v
- (Entropy(S)) adalah ukuran ketidakpastian atau impurity dalam himpunan data S

2.2 Random Forest

Random Forest adalah metode ensemble yang menggabungkan banyak decision tree untuk meningkatkan stabilitas prediksi[14]. Prediksi akhir diperoleh dengan majority voting untuk klasifikasi:

$$\hat{y} = mode\big(h_1(x), h_2(x), \dots, h_n(x)\big)$$

Dimana:
- ($h_j(x)$) adalah hasil prediksi dari pohon ke-j
- $\hat{y}$ adalah kelas hasil voting mayoritas
- n adalah jumlah pohon dalam ensemble

2.3 Support Vector Machine (SVM)

SVM mencari hyperplane optimal yang memisahkan dua kelas dengan margin maksimum [15]. Fungsi hyperplane:

$$f(x) = w^T x + b$$

Dimana:
- (W) adalah vektor bobot,
- (x) adalah vektor fitur,
- (b) adalah bias.

2.4 Logistic Regression

Logistic Regression memodelkan probabilitas kelas dengan fungsi logit:

$$P(\,y = 1 \mid x\,) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

Dimana:
- $p(Y = 1|X)$ mewakili probabilitas terjadinya peristiwa target berdasarkan kondisi predictor X.
- β_0 adalah sebagai konstanta (intercept) model.
- $\beta_1, \dots, \beta_n$ merepresentasikan koefisien regresi dari masing-masing variabel predictor.
- x_1,…..,x_n adalah nilai-nilai dari variabel prediktor yang bersesuaian.

2.5 K-Nearest Neighbors (KNN)

KNN mengklasifikasikan suatu data dengan melihat mayoritas kelas dari k tetangga terdekat, menggunakan perhitungan jarak Euclidean [16].

$$d(x, x') = \sqrt{\sum_{i=1}^{n} (x_i - x_i')^2}$$

Dimana:
- (x , x′) adalah dua vektor data yang dibandingkan
- n adalah jumlah fitur
- d(x,x′) adalah jarak Euclidean antar titik

2.6 Gradient Boosting (GB)

Gradient Boosting membangun model secara bertahap dengan menambahkan weak learner pada setiap iterasi untuk memperbaiki kesalahan model sebelumnya.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Dimana:
- $(F_m(x))$ adalah model gabungan pada iterasi ke-mmm
- $(F_{m-1}(x))$ adalah model gabungan pada iterasi sebelumnya
- $(h_m(x))$ adalah weak learner baru
- (γ_m) adalah bobot hasil minimisasi fungsi loss

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil eksperimen dari model machine learning yang digunakan dalam penelitian, serta memberikan analisis dan pembahasan terkait performa masing-masing model.

3.1. Perbandingan Model Tanpa dan Dengan SMOTE

Pada tahap awal, dilakukan evaluasi model tanpa penerapan teknik balancing. Hasilnya dapat dilihat pada Tabel 1. Secara umum, Random Forest menunjukkan performa terbaik dibandingkan dengan model lainnya, dengan akurasi 87,8% dan F1-score 0,878. Namun, performa model lain seperti Logistic Regression, KNN, dan Decision Tree masih terbatas karena adanya ketidakseimbangan kelas.

Tabel 1. Perbandingan Performa Model tanpa SMOTE

Model	Akurasi	Presisi	Recall	F1-score
Logistic Regression	0.835	0.835	0.834	0.833
K-Nearest Neighbors	0.820	0.825	0.820	0.819
Support Vector Machine	0.842	0.841	0.842	0.841
Decision Tree	0.818	0.823	0.818	0.817
Random Forest	**0.878**	**0.878**	**0.878**	**0.878**
Gradient Boosting	0.862	0.863	0.862	0.862

Setelah menerapkan SMOTE, terjadi peningkatan signifikan pada seluruh model. Yang dapat di lihat pada Tabel 2.

Tabel 2. Perbandingan Performa Model dengan SMOTE

Model	Akurasi	Presisi	Recall	F1-score
Logistic Regression	0.877	0.877	0.876	0.876
K-Nearest Neighbors	0.862	0.862	0.862	0.862
Support Vector Machine	0.882	0.882	0.882	0.882
Decision Tree	0.861	0.861	0.861	0.861
Random Forest	**0.912**	**0.912**	**0.912**	**0.912**
Gradient Boosting	0.902	0.902	0.902	0.902

Tabel 2 menunjukkan bahwa Random Forest mampu mencapai akurasi 91,2% dengan F1-score 0,912, diikuti oleh Gradient Boosting dengan akurasi 90,2%. Hal ini menunjukkan bahwa SMOTE efektif dalam mengatasi masalah ketidakseimbangan kelas, terutama pada kasus medis di mana kelas minoritas (pasien meninggal) memiliki peran penting untuk diprediksi secara tepat.

3.2. Hasil Akhir Model Terbaik pada Data Uji

Pengujian akhir dilakukan pada data uji menggunakan enam algoritma machine learning. Seperti terlihat pada Tabel 3.

Tabel 3. Hasil Akhir Model Terbaik pada Data Uji

Model	Akurasi	Presisi	Recall	F1-Score
Random Forest	0.912	0.913	0.912	0.912
Gradient boosting	0.902	0.903	0.902	0.902
Support vector machine (SVM)	0.843	0.843	0.843	0.843
K-Nearest Neighbors (KNN)	0.804	0.804	0.804	0.804
Decision Tree	0.804	0.808	0.804	0.803
Logistic Regression	0.794	0.795	0.794	0.794

Tabel 3 menunjukan bahwa model Random Forest mampu mencapai performa tertinggi dengan akurasi 91,2%, presisi 0,913, recall 0,912, dan F1-score 0,912. Hasil ini menunjukkan bahwa Random Forest tidak hanya mampu melakukan klasifikasi dengan akurasi tinggi, tetapi juga seimbang dalam mendeteksi kelas mayoritas maupun minoritas, sebagaimana ditunjukkan oleh nilai recall dan F1-score yang tinggi. Model Gradient Boosting menempati peringkat kedua dengan akurasi 90,2% dan F1-score 0,902. Support Vector Machine (SVM) memperoleh akurasi 84,3%, K-Nearest Neighbors (KNN) hanya mencapai akurasi 80,4%, Decision Tree memperoleh hasil serupa dengan KNN, yaitu akurasi 80,4% dan F1-score 0,803, Hasil terendah diperoleh oleh Logistic Regression dengan akurasi 79,4%.

3.3. Pengujian pada Dataset yang berbeda

Untuk menguji generalisasi model, dilakukan validasi menggunakan dua dataset tambahan. Hasilnya ditunjukkan pada Tabel 4. Random Forest tetap konsisten dengan performa tinggi pada kedua dataset, mencapai akurasi 100% pada Dataset Pembanding 1 dan 91,3% pada Dataset Pembanding 2.

Tabel 4. Hasil Pengujian Model Pada Dataset tambahan

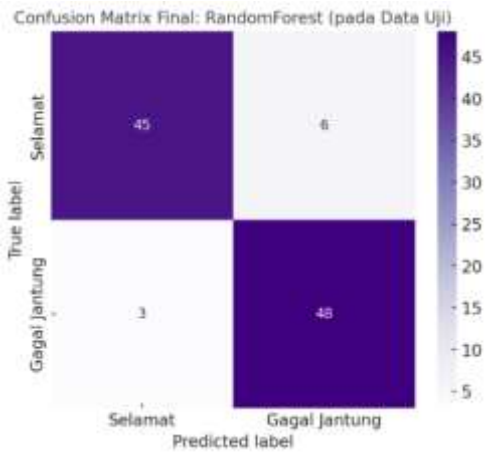
Model	heart failure clinical records dataset				Heart Disease Dataset				Heart Failure Prediction Dataset			
	Akurasi	Presisi	Recall	F1-Score	Akurasi	Presisi	Recall	F1-Score	Akurasi	Presisi	Recall	F1-Score
Random Forest	**0.912**	**0.913**	**0.912**	**0.912**	**1**	**1**	**1**	**1**	**0.913**	**0.913**	**0.913**	**0.913**
Gradient Boosting	0.902	0.903	0.902	0.902	0.957	0.957	0.957	0.957	0.898	0.899	0.898	0.898
SVM	0.843	0.843	0.843	0.843	0.907	0.907	0.907	0.907	0.89	0.89	0.89	0.89
KNN	0.804	0.804	0.804	0.804	0.852	0.853	0.853	0.852	0.882	0.883	0.882	0.882
Decision Tree	0.804	0.808	0.804	0.803	0.988	0.988	0.989	0.988	0.839	0.84	0.839	0.838
Logistic Regression	0.794	0.795	0.794	0.794	0.825	0.83	0.823	0.824	0.846	0.846	0.846	0.846

3.4. Visualisasi Confusion Matrix Masing-masing dataset

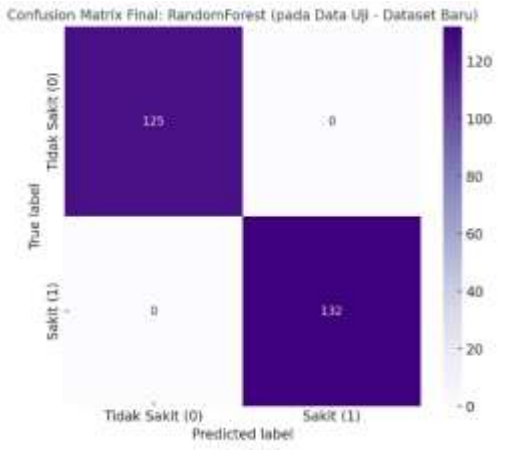
Untuk memberikan gambaran konkret terhadap kinerja model terbaik, dilakukan analisis visual melalui confusion matrix dari Random Forest. Pada Dataset awal (Heart Failure Clinical Records Dataset) yang dapat dilihat Gambar 2, confusion matrix model Random Forest menunjukkan performa yang sangat baik dalam mengenali kedua kelas target. Model berhasil memprediksi dengan tepat 45 pasien yang sebenarnya dalam kondisi Selamat (True Negative) dan 48 pasien yang sebenarnya Gagal Jantung (True Positive). Meskipun demikian, terdapat 6 pasien yang sebenarnya Selamat namun salah diprediksi sebagai 'Gagal Jantung' (False Positive).

Selanjutnya, pada Dataset Pembanding 1 (Heart Disease Dataset) yang dapat dilihat pada Gambar 3, Model berhasil memprediksi dengan tepat 125 pasien yang sebenarnya 'Tidak Sakit' (True Negative) dan 132 pasien yang sebenarnya 'Sakit' (True Positive). Hasil yang luar biasa adalah tidak adanya false positive (0 kasus), yang berarti tidak ada satu pun pasien 'Tidak Sakit' yang salah diprediksi sebagai 'Sakit'. Demikian pula, tidak ada false negative (0 kasus), menunjukkan bahwa tidak ada satu pun pasien 'Sakit' yang terlewatkan atau salah diprediksi sebagai 'Tidak Sakit'. Performa ini mengindikasikan kemampuan klasifikasi yang sangat tinggi dari Random Forest pada dataset yang seimbang.

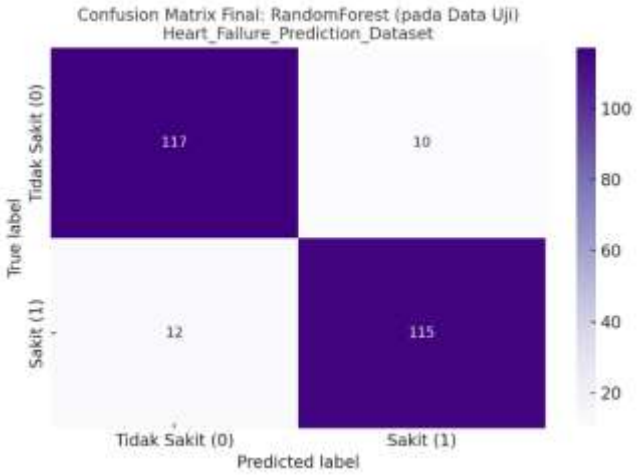
Pada Dataset Pembanding 2 (Heart Failure Prediction Dataset) yang dapat di lihat pada Gambar 4, Model berhasil memprediksi dengan tepat 117 pasien yang sebenarnya 'Tidak Sakit' (True Negative) dan 115 pasien yang sebenarnya 'Sakit' (True Positive). Namun, terdapat 10 pasien yang sebenarnya 'Tidak Sakit' namun salah diprediksi sebagai 'Sakit' (False Positive), dan 12 pasien yang kondisi aktualnya 'Sakit' namun salah diprediksi sebagai 'Tidak Sakit' (False Negative).



Gambar 2. Confusion Matrix pada Dataset awal



Gambar 3. Confusion Matrix Heart Disease Dataset



Gambar 4. Confusion Matrix Heart Failure Prediction Dataset

4. Kesimpulan

Penelitian ini menunjukkan bahwa penerapan pipeline pra-pemrosesan yang terstruktur, dengan menggabungkan seleksi fitur menggunakan Recursive Feature Elimination with Cross-validation (RFECV) dan penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE), mampu meningkatkan performa algoritma machine learning dalam klasifikasi gagal jantung. Dari hasil pengujian, model Random Forest memberikan performa terbaik dengan akurasi dan F1-score sebesar 91,2% pada dataset utama serta menunjukkan konsistensi pada dua dataset pembanding. Hasil confusion matrix juga menegaskan kemampuan Random Forest dalam menekan kesalahan klasifikasi, terutama pada kasus false negative yang krusial dalam konteks medis.

Temuan ini mengindikasikan bahwa metode ensemble, khususnya Random Forest, lebih stabil dan andal dibandingkan algoritma lain seperti Logistic Regression, KNN, atau SVM ketika dihadapkan pada data klinis yang kompleks dan tidak seimbang.

Daftar Pustaka

[1] T. A. McDonagh *et al.*, "2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," *Eur. Heart J.*, vol. 42, no. 36, pp. 3599–3726, 2021, doi: 10.1093/eurheartj/ehab368.

[2] K. Dimitrakopoulou, E. Wik, L. A. Akslen, and I. Jonassen, "Deblender: A semi-/unsupervised multi-operational computational method for complete deconvolution of expression data from heterogeneous samples," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1–17, 2018, doi: 10.1186/s12859-018-2442-5.

[3] D. Chicco and 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, pp. 1–16, 2020, doi: 10.1186/s12911-020-1023-5.

[4] J. Joseph and K. Kartheeban, "Visualizing the Full Spectrum Optimization of K-Nearest Neighbors From Data Preprocessing to Hyperparameter Tuning and K-Fold Validation for Cardiovascular Disease Prediction," *Inform.*, vol. 49, no. 2, pp. 355–374, 2025, doi: 10.31449/inf.v49i2.7774.

[5] S. Kanakala and V. Prashanthi, "Comparative analysis of heart failure prediction using machine learning models," *Int. J. Informatics Commun. Technol.*, vol. 13, no. 2, pp. 297–305, 2024, doi: 10.11591/ijict.v13i2.pp297-305.

[6] A. Masitha, M. K. Biddinika, and H. Herman, "K Value Effect on Accuracy Using the K-NN for Heart Failure Dataset," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 3, pp. 593–604, 2023, doi: 10.30812/matrik.v22i3.2984.

[7] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor Dan Logistic Regression," *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 493, 2022, doi: 10.37600/tekinkom.v5i2.698.

[8] L. Anggraini *et al.*, "Evaluasi Logistic Regression dan Neural Network pada Klasifikasi Gagal Jantung Berbasis Threshold," *J. Tek. Inform. Unika ST. Thomas*, vol. 10, pp. 14–23, 2025.

[9] R. Ridwan, H. H. Handayani, S. A. P. Lestari, and ..., "Evaluasi Kinerja Algoritma Random Forest Dan Gradient Boosting Untuk Klasifikasi Penyakit Jantung," *J. Komtika …*, vol. 9, no. 1, pp. 112–124, 2025, [Online]. Available: https://journal.unimma.ac.id/index.php/komtika/article/view/13450%0Ahttps://journal.unimma.ac.id/index.php/komtika/article/download/13450/5846/

[10] A. A. Putri Lo and V. J. E. Tjioe, "Penerapan Model CRISP-DM untuk Prediksi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbor dan Logistic regression," *Pros. SENAM 2024 Semin. Nas. Sist. Inf. Inform. Univ. Ma Chung*, vol. 4, pp. 48–57, 2024, [Online]. Available: https://ocs.machung.ac.id/index.php/seminarnasionalmachung/article/view/452

[11] N. Jannath and M. J. S. Masud, Md AbdulHossain, "Improving Classification Using SMOTE on Imbalanced Heart Failure Data," *J. Patuakhali Sei. Tech. Uni*, vol. 12, no. 1 & 2, pp. 29–43, 2022.

[12] P. tua Sinaga, S. S. Purba, D. Wiranto, O. J. Maharja, and E. Indra, "Exploratory Data Analysis of Clinical Heart Failure Using a Support Vector Machine," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 7, no. 1, pp. 142–154, 2023, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v7i1.4100.

[13] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," Healthc. Anal., vol. 3, no. December 2022, p. 100130, 2023, doi: 10.1016/j.health.2022.100130

[14] N. Dwi, A. Nur, J. Tri, and K. Arai, "Ensemble learning approaches for predicting heart failure outcomes : A comparative analysis of feedforward neural networks , random forest , and XGBoost," *Appl. Eng. Technol.*, vol. 3, no. 3, pp. 173–184, 2024.

[15] Y. Tampil, H. Komaliq, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *d'CARTESIAN*, vol. 6, no. 2, p. 56, 2017, doi: 10.35799/dc.6.2.2017.17023

[16] H. A. Taher and A. M. Abdulazeez, "Machine Learning Approaches for Heart Disease Detection: A Comprehensive Review," *Int. J. Res. Appl. Technol.*, vol. 3, no. 2, pp. 267–282, 2023