

KOMPARASI ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI RISIKO SERANGAN JANTUNG DI INDONESIA

Rizal Maulana¹⁾, Ahmad Al Kaafi²⁾, Rizka Dahlia³⁾

^{1,2,3}Universitas Bina Sarana Informatika, Jakarta, Indonesia

Corresponding Author: rizalmln2004@gmail.com

Article Info

Article history:

Received: Aug 09, 2026

Revised: Aug 18, 2026

Accepted: Aug 20, 2026

Published: Oct 15, 2026

Keywords:

Serangan Jantung,
Machine Learning,
Klasifikasi,
Feature Selection,
Komparasi.

ABSTRACT

Serangan jantung merupakan salah satu penyebab kematian tertinggi di Indonesia sehingga diperlukan upaya deteksi dini untuk mengidentifikasi individu yang berisiko mengalami kondisi tersebut. Penelitian ini bertujuan membandingkan performa algoritma Decision Tree, Random Forest, dan Gradient Boosting dalam klasifikasi risiko serangan jantung di Indonesia, serta menganalisis pengaruh feature selection terhadap performa model. Penelitian menggunakan dataset Heart Attack Prediction in Indonesia yang diperoleh dari Kaggle, yang terdiri atas 158.355 data individu dan 28 atribut. Data tersebut mencakup karakteristik individu dan faktor yang berkaitan dengan risiko serangan jantung, seperti usia, jenis kelamin, hipertensi, diabetes, kadar kolesterol, obesitas, kebiasaan merokok, aktivitas fisik, tekanan darah, serta riwayat penyakit jantung sebelumnya. Tahapan penelitian meliputi data cleaning, transformasi data menggunakan one-hot encoding, feature selection menggunakan metode Gain Ratio, serta pembagian data menjadi 80% data latih dan 20% data uji. Performa model dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Gradient Boosting memperoleh akurasi tertinggi sebesar 73,6%, diikuti Decision Tree sebesar 72,2% dan Random Forest sebesar 71,2%. Feature selection menghasilkan 9 atribut terpilih tanpa memberikan peningkatan performa yang berarti. Dengan demikian, Gradient Boosting merupakan algoritma dengan performa terbaik dalam klasifikasi risiko serangan jantung pada penelitian ini.



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. INTRODUCTION

Sebagai organ vital, jantung bertanggung jawab menjaga kelancaran sirkulasi darah ke seluruh tubuh melalui sistem pembuluh darah, sehingga sangat penting bagi keberlangsungan hidup manusia. Namun, adanya penyakit jantung dapat mengganggu fungsi organ tersebut dan menimbulkan dampak serius terhadap Kesehatan [1].

Menurut data Survei Kesehatan Indonesia (SKI) tahun 2023, penyakit jantung masih menjadi isu kesehatan yang memerlukan perhatian serius di Indonesia, dengan prevalensi nasional berdasarkan diagnosis dokter tercatat sebesar 0,85% (sekitar 85 dari 10.000 penduduk). Tingginya angka tersebut menunjukkan bahwa penyakit jantung masih menjadi permasalahan kesehatan yang serius di Indonesia. Selain itu, proses diagnosis penyakit jantung umumnya memerlukan berbagai pemeriksaan medis yang menghasilkan data kesehatan pasien, seperti riwayat penyakit jantung, kadar kolesterol, dan

indikator lainnya. Namun, pemanfaatan data tersebut masih belum optimal dalam membantu proses identifikasi risiko secara cepat dan akurat [2].

Kemajuan pada teknologi machine learning telah mendorong pemanfaatan berbagai pendekatan berbasis data untuk mendukung analisis kesehatan. Dengan memanfaatkan data yang tersedia, sistem dapat mempelajari pola serta melakukan prediksi secara lebih cepat dan akurat. Penelitian oleh [3] menunjukkan bahwa "Penerapan machine learning terbukti mampu mengidentifikasi potensi serangan jantung secara efektif sehingga mendukung proses deteksi dini dan pengambilan keputusan medis."

Pada bidang machine learning, klasifikasi merupakan metode yang banyak diterapkan, yaitu teknik analisis data untuk memprediksi label atau kategori tertentu berdasarkan pola yang terbentuk [4]. Untuk mengidentifikasi faktor yang paling

mempengaruhi terjadinya serangan jantung, metode klasifikasi merupakan pendekatan yang tepat.

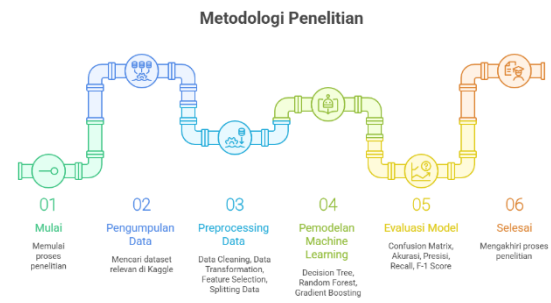
Selain itu, dalam proses pengolahan data, tidak semua variabel memiliki pengaruh yang signifikan terhadap hasil klasifikasi. Untuk mengatasi hal tersebut, penerapan feature selection menjadi solusi dalam menyaring atribut yang paling berdampak terhadap risiko serangan jantung. Proses ini bermanfaat untuk mengoptimalkan efisiensi pengolahan data, menyederhanakan kompleksitas atribut, serta membantu model dalam memproses data yang lebih relevan. Dengan demikian, penerapan seleksi fitur menjadi salah satu langkah penting dalam proses klasifikasi machine learning [5].

Meskipun berbagai penelitian telah menerapkan algoritma machine learning untuk klasifikasi penyakit jantung, hasil performa setiap algoritma masih menunjukkan variasi sehingga belum dapat ditentukan algoritma yang paling sesuai untuk digunakan pada dataset risiko serangan jantung di Indonesia. Decision Tree merupakan algoritma yang membentuk struktur pohon untuk menentukan kelas berdasarkan atribut yang digunakan dalam proses pengambilan keputusan [6]. Random Forest merupakan algoritma ensemble yang menggabungkan beberapa Decision Tree dengan pemilihan data dan atribut secara acak, kemudian menentukan hasil klasifikasi berdasarkan suara mayoritas [7]. Sementara itu, Gradient Boosting merupakan algoritma ensemble yang membangun Decision Tree secara bertahap dengan tujuan memperbaiki kesalahan prediksi dari model sebelumnya [8]. Selain itu, penerapan feature selection sebagai upaya mengidentifikasi atribut yang paling relevan masih perlu dikaji lebih lanjut terhadap performa model klasifikasi. Berdasarkan permasalahan tersebut, penelitian ini bertujuan membandingkan performa algoritma Decision Tree, Random Forest, dan Gradient Boosting dalam klasifikasi risiko serangan jantung di Indonesia dengan menerapkan feature selection menggunakan metode Gain Ratio. Performa masing-masing algoritma kemudian dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score untuk menentukan algoritma yang memberikan hasil klasifikasi terbaik.

2. MATERIALS AND METHODS

Untuk memberikan gambaran yang jelas mengenai tahapan penelitian, proses penelitian dilakukan secara sistematis mulai dari tahap awal hingga tahap akhir. Setiap tahapan dirancang untuk memastikan data yang digunakan telah diproses dengan baik sehingga hasil klasifikasi dan perbandingan performa algoritma dapat dilakukan secara objektif. Alur metode penelitian tersebut ditunjukkan pada Gambar 1.

2.1. Pengumpulan Data



Gambar 1. Metode Penelitian

Penelitian ini menggunakan dataset sekunder yang diperoleh dari repositori Kaggle berjudul Heart Attack Prediction in Indonesia [9]. Dataset tersebut terdiri atas 158.355 data individu dengan 28 atribut yang merepresentasikan kondisi kesehatan, gaya hidup, dan faktor lingkungan, seperti usia, kadar kolesterol, riwayat hipertensi, diabetes, kebiasaan merokok, serta riwayat penyakit jantung sebelumnya. Variabel target pada dataset ini adalah heart_attack, yang terdiri atas dua kelas, yaitu 0 (tidak mengalami serangan jantung) dan 1 (mengalami serangan jantung).

2.2. Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan dataset sebelum proses pemodelan machine learning. Tahap ini bertujuan untuk memastikan data yang digunakan memiliki kualitas yang baik sehingga dapat menghasilkan model klasifikasi yang optimal [10]. Proses preprocessing meliputi pengecekan missing value, transformasi data kategorikal menggunakan teknik one-hot encoding, seleksi fitur menggunakan metode Gain Ratio, serta pembagian data menjadi data latih dan data uji dengan perbandingan 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi performa model.

2.3. Pemodelan Machine Learning

Tahap pemodelan machine learning dilakukan untuk membangun model klasifikasi risiko serangan jantung berdasarkan data latih yang telah melalui proses preprocessing. Pada penelitian ini digunakan tiga algoritma, yaitu Decision Tree, Random Forest, dan Gradient Boosting, yang kemudian dibandingkan berdasarkan performa klasifikasinya.

Decision Tree merupakan algoritma machine learning yang merepresentasikan proses pengambilan keputusan dalam bentuk struktur pohon [11]. Random Forest merupakan algoritma ensemble yang membangun beberapa Decision Tree dan menentukan hasil klasifikasi berdasarkan suara mayoritas (majority

voting) dari seluruh pohon [12]. Sementara itu, Gradient Boosting merupakan algoritma ensemble yang membangun Decision Tree secara bertahap untuk memperbaiki kesalahan prediksi dari model sebelumnya sehingga menghasilkan performa klasifikasi yang lebih baik [13].

2.3. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa setiap algoritma dalam mengklasifikasikan risiko serangan jantung menggunakan data uji. Pada penelitian ini, evaluasi dilakukan menggunakan confusion matrix, yaitu metode evaluasi yang menyajikan jumlah prediksi benar maupun salah berdasarkan kelas aktual dan kelas hasil prediksi [14]. Berdasarkan confusion matrix, performa model diukur menggunakan metrik accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan, precision menunjukkan ketepatan model dalam memprediksi kelas positif, recall mengukur kemampuan model dalam mengidentifikasi seluruh data positif, sedangkan F1-score merupakan rata-rata harmonis antara precision dan recall yang memberikan gambaran keseimbangan performa model [15].

3. RESULTS AND DISCUSSION

Bagian ini menyajikan hasil penelitian mengenai klasifikasi risiko serangan jantung di Indonesia menggunakan algoritma Decision Tree, Random Forest, dan Gradient Boosting. Hasil yang diperoleh meliputi perbandingan performa masing-masing algoritma serta evaluasi model. Pembahasan dilakukan untuk menganalisis hasil pengolahan data dan menentukan algoritma yang memberikan performa terbaik dalam klasifikasi risiko serangan jantung

3.1. Hasil Pengumpulan Data

Dataset yang digunakan dalam penelitian ini bersumber dari Heart Attack Prediction in Indonesia dengan total 158.355 data dan 28 atribut. Atribut yang tersedia terdiri atas data kategorikal dan numerikal yang merepresentasikan kondisi kesehatan, gaya hidup, serta faktor lingkungan yang berkaitan dengan risiko serangan jantung. Rincian atribut dataset disajikan pada Tabel 1.

Tabel 1. Identifikasi Atribut dan Tipe Data

No	Atribut	Tipe
1	age	numerikal
2	gender	kategorikal
3	region	kategorikal
4	income_level	kategorikal

5	hypertension	kategorikal
6	diabetes	kategorikal
7	cholesterol_level	numerikal
8	obesity	kategorikal
9	waist_circumference	numerikal
10	family_history	kategorikal
11	smoking_status	kategorikal
12	alcohol_consumption	kategorikal
13	physical_activity	kategorikal
14	dietary_habits	kategorikal
15	air_pollution_exposure	kategorikal
16	stress_level	kategorikal
17	sleep_hours	numerikal
18	blood_pressure_systolic	numerikal
19	blood_pressure_diastolic	numerikal
20	fasting_blood_sugar	numerikal
21	cholesterol_hdl	numerikal
22	cholesterol_ldl	numerikal
23	triglycerides	numerikal
24	EKG_results	kategorikal
25	previous_heart_disease	kategorikal
26	medication_usage	kategorikal
27	participated_in_free_screening	kategorikal
28	heart_attack	kategorikal

3.2. Hasil Preprocessing Data

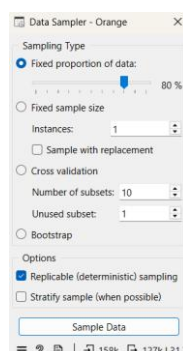
Tahap preprocessing dilakukan untuk mempersiapkan dataset sebelum proses pemodelan machine learning. Pada tahap data cleaning, dilakukan pengecekan terhadap missing value untuk memastikan kualitas dan kelengkapan data yang digunakan. Hasil pengecekan menunjukkan bahwa dataset yang terdiri atas 158.355 data dan 28 atribut tidak memiliki missing value, sehingga tidak diperlukan proses penanganan atau imputasi terhadap data yang hilang. Selanjutnya, pada tahap data transformation, atribut kategorikal ditransformasikan ke dalam bentuk numerik menggunakan teknik one-hot encoding. Setiap kategori direpresentasikan dalam bentuk nilai biner 0 dan 1 sehingga dapat digunakan dalam proses pemodelan. Tahap berikutnya adalah feature selection menggunakan metode Gain Ratio untuk mengidentifikasi atribut yang paling relevan terhadap proses klasifikasi risiko serangan jantung. Berdasarkan hasil perhitungan, previous_heart_disease memperoleh nilai Gain Ratio

sebesar 0,074, diikuti oleh hypertension sebesar 0,059 dan diabetes sebesar 0,037. Atribut lain yang turut dipilih meliputi obesity, smoking_status, age, cholesterol_level, waist_circumference, dan fasting_blood_sugar. Berdasarkan hasil feature selection, dipilih 9 atribut yang memiliki nilai Gain Ratio tertinggi untuk digunakan dalam proses pemodelan. Hasil feature selection menggunakan metode Gain Ratio ditampilkan pada Tabel II.

Tabel 2. Hasil Feature Selection dengan Gain Ratio

No	Atribut	Gain Ratio
1	previous_heart_disease=0	0.074
	previous_heart_disease=1	0.074
2	hypertension=0	0.059
	hypertension=1	0.059
3	diabetes=0	0.037
	diabetes=1	0.037
4	obesity=0	0.026
	obesity=1	0.026
5	smoking_status = current	0.026
	smoking_status = never	0.005
	smoking_status = past	0.002
6	age	0.005
7	cholesterol_level	0.004
8	waist_circumference	0.001
9	fasting_blood_sugar	0.001

Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Sebanyak 126.684 data (80%) digunakan sebagai data latih untuk membangun model, sedangkan 31.671 data (20%) digunakan sebagai data uji untuk mengevaluasi performa model. Konfigurasi pembagian data ditunjukkan pada Gambar 2.



Gambar 2. Konfigurasi Splitting Data

3.3. Hasil Komparasi Performa Model

Pengujian dilakukan menggunakan algoritma Decision Tree, Random Forest, dan Gradient Boosting untuk membandingkan performa masing-masing model dalam klasifikasi risiko serangan jantung. Evaluasi performa dilakukan berdasarkan nilai accuracy, precision, recall, dan F1-score. Hasil perbandingan performa ketiga algoritma disajikan pada Tabel III.

Tabel 3. Hasil Pengujian Model

Model	CA	Prec	Recall	F1
Decision Tree	0.722	0.718	0.722	0.718
Random Forest	0.712	0.709	0.712	0.709
Gradient Boosting	0.736	0.733	0.736	0.734

Berdasarkan Tabel III, Gradient Boosting memperoleh performa tertinggi dengan accuracy sebesar 73,6%, diikuti Decision Tree sebesar 72,2% dan Random Forest sebesar 71,2%. Gradient Boosting juga memperoleh nilai precision, recall, dan F1-score tertinggi dibandingkan kedua algoritma lainnya, sehingga memberikan performa klasifikasi terbaik pada dataset yang digunakan.

3.4. Evaluasi Model

Evaluasi performa model dilakukan untuk mengetahui kemampuan masing-masing algoritma dalam mengklasifikasikan risiko serangan jantung berdasarkan data uji. Evaluasi dilakukan menggunakan confusion matrix yang menunjukkan jumlah prediksi benar dan salah pada masing-masing kelas, yaitu True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP).

3.4.1. Decision Tree

		Predicted		
		0	1	Σ
Actual	0	15452	3624	19076
	1	5192	7403	12595
Σ		20644	11027	31671

Gambar 3. Confusion Matrix Decision Tree

Berdasarkan hasil confusion matrix, algoritma Decision Tree mampu mengklasifikasikan sebanyak 15.452 data kelas negatif dengan benar (True Negative) dan 7.403 data kelas positif dengan benar (True Positive). Model juga menghasilkan 3.624 False Positive, yaitu data negatif yang diprediksi sebagai positif, serta 5.192 False Negative, yaitu data positif yang diprediksi sebagai negatif. Hasil tersebut

menunjukkan bahwa Decision Tree mampu mengklasifikasikan kedua kelas, meskipun masih terdapat kesalahan dalam mengidentifikasi data positif.

3.4.2. Random Forest

		Predicted		
		0	1	Σ
Actual	0	15060	4016	19076
	1	5108	7487	12595
Σ		20168	11503	31671

Gambar 4. Confusion Matrix Random Forest

Berdasarkan hasil confusion matrix, Random Forest menghasilkan 15.060 True Negative dan 7.487 True Positive, dengan 4.016 False Positive dan 5.108 False Negative. Hasil tersebut menunjukkan bahwa Random Forest mampu mengenali data positif dengan lebih baik dibandingkan Decision Tree berdasarkan jumlah True Positive, tetapi menghasilkan jumlah False Positive yang lebih tinggi. Selain itu, terdapat penurunan performa pada data uji yang tercermin dari nilai accuracy sebesar 71,2%.

3.4.2. Gradient Boosting

		Predicted		
		0	1	Σ
Actual	0	15287	3789	19076
	1	4587	8008	12595
Σ		19874	11797	31671

Gambar 5. Confusion Matrix Gradient Boosting

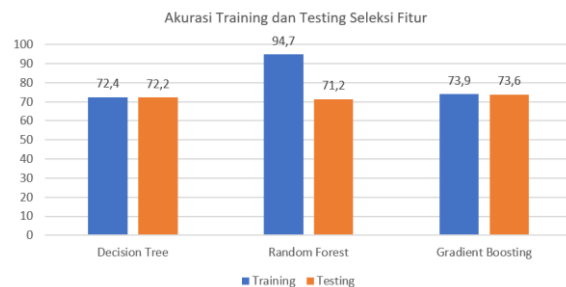
Pada algoritma Gradient Boosting, sebanyak 15.287 data kelas negatif berhasil diklasifikasikan dengan benar dan 8.008 data kelas positif berhasil diprediksi secara tepat. Model menghasilkan 3.789 False Positive dan 4.587 False Negative. Dibandingkan Decision Tree dan Random Forest, Gradient Boosting memiliki jumlah True Positive tertinggi dan False Negative terendah, sehingga menunjukkan kemampuan yang lebih baik dalam mengidentifikasi data positif. Hasil tersebut sejalan dengan nilai accuracy sebesar 73,6%, yang merupakan nilai tertinggi di antara ketiga algoritma.

Berdasarkan hasil confusion matrix, Gradient Boosting menunjukkan performa terbaik dalam mengidentifikasi kelas positif dengan jumlah True Positive tertinggi sebesar 8.008 dan False Negative terendah sebesar 4.587. Hasil ini sejalan dengan

perbandingan performa model, di mana Gradient Boosting memperoleh accuracy sebesar 73,6%, diikuti Decision Tree sebesar 72,2% dan Random Forest sebesar 71,2%.

3.5. Analisis Overfitting & Stabilitas Model

Analisis overfitting dilakukan dengan membandingkan akurasi training dan testing untuk menilai kemampuan model dalam melakukan generalisasi terhadap data baru. Hasil perbandingan akurasi ketiga algoritma ditunjukkan pada Gambar 6.



Gambar 6. Grafik Perbandingan Akurasi Training dan Testing

Berdasarkan Gambar 6, algoritma Decision Tree dan Gradient Boosting menunjukkan performa yang stabil. Decision Tree memperoleh akurasi training sebesar 72,4% dan testing sebesar 72,2%, dengan selisih hanya 0,2%. Sementara itu, Gradient Boosting memperoleh akurasi training sebesar 73,9% dan testing sebesar 73,6%, dengan selisih 0,3%. Selisih yang kecil pada kedua algoritma tersebut menunjukkan bahwa model mampu melakukan generalisasi dengan baik terhadap data baru sehingga tidak mengindikasikan terjadinya overfitting. Sebaliknya, algoritma Random Forest menunjukkan indikasi overfitting. Model ini memperoleh akurasi training sebesar 94,7%, sedangkan akurasi testing hanya mencapai 71,2%, sehingga terdapat selisih sebesar 23,5%. Perbedaan yang cukup besar antara akurasi training dan testing menunjukkan bahwa model memiliki performa yang sangat baik pada data latih, namun kemampuannya menurun ketika diuji menggunakan data baru. Hal tersebut mengindikasikan bahwa model belum mampu melakukan generalisasi secara optimal terhadap data yang belum pernah dipelajari. Berdasarkan hasil analisis tersebut, Gradient Boosting merupakan algoritma yang paling stabil karena menghasilkan selisih akurasi training dan testing yang sangat kecil serta memperoleh akurasi testing tertinggi sebesar 73,6%. Decision Tree juga menunjukkan stabilitas yang baik dengan selisih akurasi hanya 0,2%, sedangkan Random Forest menunjukkan indikasi overfitting sehingga performanya pada data testing menjadi lebih rendah dibandingkan kedua algoritma lainnya meskipun memiliki akurasi training yang paling tinggi.

4. CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Gradient Boosting memperoleh performa terbaik dalam klasifikasi risiko serangan jantung di Indonesia dengan nilai accuracy sebesar 73,6%, diikuti oleh Decision Tree sebesar 72,2% dan Random Forest sebesar 71,2%. Meskipun demikian, perbedaan performa ketiga algoritma relatif kecil. Hasil analisis overfitting menunjukkan bahwa Gradient Boosting dan Decision Tree memiliki stabilitas model yang baik karena menghasilkan selisih akurasi training dan testing yang kecil, sedangkan Random Forest menunjukkan indikasi overfitting dengan selisih akurasi training dan testing yang cukup besar. Selain itu, proses feature selection menggunakan metode Gain Ratio menghasilkan sembilan atribut yang paling relevan dalam klasifikasi risiko serangan jantung. Namun, penggunaan sembilan atribut tersebut tidak memberikan peningkatan performa yang berarti dibandingkan penggunaan seluruh atribut.

Penelitian selanjutnya dapat menggunakan dataset yang lebih beragam dan representatif agar hasil klasifikasi mampu menggambarkan kondisi risiko serangan jantung di Indonesia secara lebih baik. Selain itu, optimasi hyperparameter, khususnya pada algoritma Random Forest, dapat dilakukan untuk mengurangi indikasi overfitting sehingga kemampuan generalisasi model terhadap data baru dapat ditingkatkan.

ACKNOWLEDGEMENTS

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bantuan, dan motivasi selama proses penelitian hingga penyusunan artikel ini.

REFERENCES

- [1] M. Cahaya Rani, F. Diba Azkia, R. Arrova Dewi, M. Wahyudi, A. Surya Budiman, and C. Author, "PERBANDINGAN ALGORITMA RANDOM FOREST, NAIVE BAYES, DAN NEURAL NETWORK DALAM KLASIFIKASI PENYAKIT JANTUNG," Bulan juni, 2025.
- [2] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *JURNAL INFORMATIK Edisi ke*, vol. 18, p. 2022, Dec. 2022.
- [3] G. Ferry Mandias and I. J. Manoppo, "Penerapan Model Machine Learning untuk Memprediksi Serangan Jantung Dini," *Jurnal Ilmiah Matrik*, vol. 27, no. 2, Aug. 2025.
- [4] C. A. Bahri *et al.*, "ANALISIS FAKTOR RISIKO PEMICU SERANGAN JANTUNG DI INDONESIA, MENGGUNAKAN METODE KLASIFIKASI (DECISION TREE, NAIVE BAYES, DAN RANDOM FOREST)," 2025.
- [5] S. Narulita and N. Adi, "Feature Selection Information Gain pada Klasifikasi Pasien Penyakit Jantung (Heart Disease)," vol. 4, no. 1, Jun. 2024.
- [6] A. Irma Purnamasari and T. Suprapti, "PENERAPAN ALGORITMA DECISION TREE DALAM KLASIFIKASI PENYAKIT STROKE OTAK," Cirebon, Jun. 2024.
- [7] R. Ridwan, H. H. Handayani, S. A. P. Lestari, and Y. Cahyana, "Evaluasi Kinerja Algoritma Random Forest Dan Gradient Boosting Untuk Klasifikasi Penyakit Jantung," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 9, no. 1, pp. 112–124, Jun. 2025, doi: 10.31603/komtika.v9i1.13450.
- [8] Hidayat Jose Julian and Saputra Muhammad Randika, "Evaluasi Kinerja Algoritma Ensemble Learning Pada Klasifikasi Penyakit Diabetes Berbasis Boosting Method," *Jurnal Surya Informatika*, vol. 16, pp. 71–80, May 2026.
- [9] Ankush Panday, "Heart Attack Prediction in Indonesia," Kaggle.
- [10] E. F. Laili, Z. Alawi, R. Rohmah, and A. Barata, "KOMPARASI ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI SERANGAN JANTUNG," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025, [Online]. Available: <https://www.kaggle.com/datasets/thxogg/heart->
- [11] T. Indriyani *et al.*, "Metode Decision Tree C4.5 untuk Klasifikasi penyakit Jantung," 2024.
- [12] I. G. A. G. A. K. I Made Krisna Dwipa Jaya, "Perbandingan Random Forest, Decision Tree, Gradient Boosting, Logistic Regression untuk Klasifikasi Penyakit Jantung," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, Nov. 2023.
- [13] J. P. Anggraini, Chaya Gladys Zhafirah A, and A. Desiani, "Perbandingan Algoritma Random Forest dan Extreme Gradient Boosting (XGBoost) dalam Klasifikasi Penyakit Gagal Jantung," *Komputika : Jurnal Sistem Komputer*, vol. 14, no. 2, pp. 149–157, Nov. 2025, doi: 10.34010/komputika.v14i2.16618.
- [14] F. Reynaldi Valerian, M. Syarif, D. Abdul Fatah, J. Raya Telang, K. Kamal, and J. Timur, "KLASIFIKASI TINGKAT OBESITAS MENGGUNAKAN METODE GBM DAN CONFUSION MATRIX," 2025.
- [15] D. Galih Pradana, M. L. Alghifari, M. Farhan Juna, and S. Dwisiwi Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 2, pp. 55–60, 2022.