



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1159–1167

Evaluasi Performa Algoritma C4.5 untuk Klasifikasi Gagal Jantung Berbasis *Heart Failure Clinical Records* dengan RapidMiner

Nasrul Isrofan, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister>

DOI : <https://doi.org/10.32672/mister.v3i1.4042>

How to Cite this Article

APA : Isrofan, N., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Evaluasi Performa Algoritma C4.5 untuk Klasifikasi Gagal Jantung Berbasis Heart Failure Clinical Records dengan RapidMiner. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1). <https://doi.org/10.32672/mister.v3i1.4042>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



9

773032

710001

9

773032

601002

Evaluasi Performa Algoritma C4.5 untuk Klasifikasi Gagal Jantung Berbasis *Heart Failure Clinical Records* dengan RapidMiner

Nasrul Isrofan¹, Hasbi Firmansyah², Wahyu Asriyani³

¹S1 Informatika Universitas Pancasakti Tegal

²Informatika, Fakultas Teknik, Universitas Pancasakti Tegal, Indonesia

³Pendidikan Bahasa Dan Sastra Indonesia, Fakultas Keguruan Dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Indonesia

*Email nasrultampan01091@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani409@gmail.com

Diterima: 14-12-2025

| Disetujui: 24-12-2025

| Diterbitkan: 26-12-2025

ABSTRACT

Heart failure is one of the cardiovascular diseases with a high mortality rate, requiring data analysis methods that can support accurate decision-making. The problem addressed in this study is how to evaluate the performance of classification algorithms in predicting heart failure patient conditions based on clinical data. This study aims to evaluate the performance of the C4.5 algorithm in classifying heart failure patients using the Heart Failure Clinical Records dataset. The research method applies a data mining classification approach, where data processing and model training are conducted using RapidMiner software. The research stages include data preprocessing, data splitting, C4.5 decision tree model construction, and performance evaluation using a confusion matrix. The results show that the C4.5 algorithm is able to classify heart failure patient data with good accuracy and produces an interpretable decision tree model. It can be concluded that the C4.5 algorithm is an effective method for heart failure classification based on clinical data.

Keywords: c4.5, classification, data mining, heart failure, rapidminer

ABSTRAK

Gagal jantung merupakan salah satu penyakit kardiovaskular dengan tingkat mortalitas yang tinggi sehingga diperlukan metode analisis data yang mampu membantu proses pengambilan keputusan secara akurat. Permasalahan dalam penelitian ini adalah bagaimana mengevaluasi performa algoritma klasifikasi dalam memprediksi kondisi pasien gagal jantung berdasarkan data klinis. Penelitian ini bertujuan untuk mengevaluasi performa algoritma C4.5 dalam melakukan klasifikasi pasien gagal jantung menggunakan Heart Failure Clinical Records. Metode yang digunakan adalah data mining dengan pendekatan klasifikasi, dimana proses pengolahan data dan pelatihan model dilakukan menggunakan perangkat lunak RapidMiner. Tahapan penelitian meliputi preprocessing data, pembagian data latih dan data uji, pembangunan model decision tree C4.5, serta evaluasi performa menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan data pasien gagal jantung dengan tingkat akurasi yang baik, serta menghasilkan model yang mudah dipahami melalui struktur pohon keputusan. Kesimpulan dari penelitian ini adalah algoritma C4.5 dapat digunakan sebagai metode yang efektif dalam klasifikasi data gagal jantung berbasis data klinis.

Kata kunci: c4.5, data mining, gagal jantung, klasifikasi, rapidminer

PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia dan menjadi masalah kesehatan yang signifikan, khususnya di negara berkembang. Penyakit ini sering kali sulit dideteksi pada tahap awal karena gejalanya yang tidak selalu spesifik, sehingga diperlukan pendekatan yang mampu membantu tenaga medis dalam melakukan diagnosis dan prediksi secara lebih akurat dan cepat. Perkembangan teknologi informasi, khususnya di bidang kecerdasan buatan, membuka peluang besar dalam pemanfaatan data medis untuk mendukung pengambilan keputusan klinis secara objektif dan berbasis data [1].

Machine learning telah banyak digunakan dalam bidang kesehatan untuk melakukan klasifikasi, prediksi, dan analisis penyakit jantung. Berbagai penelitian menunjukkan bahwa algoritma machine learning mampu memberikan performa yang baik dalam memprediksi kelangsungan hidup pasien maupun mendeteksi risiko penyakit jantung berdasarkan data klinis [1], [3]. Dataset Heart Failure Clinical Records yang disediakan oleh UCI Machine Learning Repository menjadi salah satu dataset yang paling sering digunakan karena memiliki atribut klinis yang relevan dan bersifat terbuka untuk penelitian [2].

Di antara berbagai algoritma machine learning, metode decision tree menjadi salah satu pendekatan yang populer karena kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan. Algoritma decision tree seperti C4.5 dan CART telah banyak diterapkan dalam klasifikasi penyakit jantung dan menunjukkan hasil yang kompetitif dibandingkan metode lainnya [6], [7]. Selain memiliki tingkat akurasi yang baik, decision tree juga memberikan visualisasi berupa struktur pohon keputusan yang dapat membantu dalam memahami hubungan antar atribut dan proses pengambilan keputusan secara sistematis.

Sejumlah penelitian terdahulu telah melakukan perbandingan berbagai algoritma klasifikasi untuk prediksi penyakit jantung, termasuk support vector machine, random forest, naive Bayes, dan deep learning [4], [8], [10]. Hasil penelitian tersebut menunjukkan bahwa tidak terdapat satu algoritma yang selalu unggul dalam semua kondisi, sehingga pemilihan metode perlu disesuaikan dengan karakteristik data dan tujuan analisis. Beberapa studi juga menekankan pentingnya seleksi fitur dan interpretabilitas model dalam konteks aplikasi medis, agar hasil yang diperoleh dapat digunakan secara praktis oleh tenaga kesehatan [9], [11].

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada penerapan algoritma decision tree untuk klasifikasi penyakit jantung menggunakan data klinis pasien. Penelitian ini diharapkan dapat memberikan gambaran kinerja algoritma decision tree serta menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga mudah dipahami. Selain itu, visualisasi pohon keputusan yang dihasilkan diharapkan dapat memberikan wawasan tambahan terkait faktor-faktor yang berpengaruh terhadap penyakit jantung, sehingga dapat mendukung proses pengambilan keputusan dalam bidang kesehatan secara lebih efektif dan transparan.

METODE PENELITIAN

Penelitian ini memfokuskan pada permasalahan klasifikasi kondisi pasien gagal jantung berdasarkan data klinis. Dataset yang digunakan adalah Heart Failure Clinical Records yang terdiri dari sejumlah atribut klinis dan satu atribut target. Secara umum, permasalahan yang diteliti dapat diformulasikan sebagai permasalahan klasifikasi biner.

Misalkan diberikan sebuah dataset

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

di mana $x_i = (a_1, a_2, \dots, a_m)$ merepresentasikan vektor atribut klinis pasien ke-i, dan $y_i \in \{0,1\}$ merupakan kelas target yang menyatakan kondisi pasien (tidak meninggal dan meninggal).

Tujuan penelitian ini adalah membangun suatu model klasifikasi yang mampu memetakan data klinis pasien ke dalam kelas yang sesuai dengan tingkat kesalahan seminimal mungkin. Evaluasi performa model dilakukan menggunakan metrik akurasi yang diperoleh dari confusion matrix.

Metode yang Diusulkan

Metode yang diusulkan dalam penelitian ini menggunakan pendekatan data mining dengan teknik klasifikasi untuk memprediksi kondisi pasien gagal jantung berdasarkan data klinis. Pendekatan klasifikasi dipilih karena mampu mengelompokkan data pasien ke dalam kelas tertentu berdasarkan pola yang dipelajari dari data historis [1], [4]. Algoritma yang digunakan adalah C4.5, yaitu salah satu algoritma decision tree yang banyak diterapkan dalam bidang medis karena menghasilkan model yang mudah dipahami dan bersifat interpretatif [6], [7].

Proses klasifikasi dilakukan melalui beberapa tahapan utama, meliputi preprocessing data, pembagian data menjadi data latih dan data uji, pembangunan model klasifikasi, serta evaluasi performa model. Seluruh tahapan tersebut diimplementasikan menggunakan perangkat lunak RapidMiner, yang telah banyak digunakan dalam penelitian data mining karena menyediakan lingkungan visual yang memudahkan proses pemodelan dan evaluasi algoritma [10], [15].

Penggunaan algoritma C4.5 dalam penelitian ini diharapkan dapat menghasilkan model klasifikasi yang tidak hanya memiliki performa yang baik, tetapi juga mampu memberikan penjelasan yang jelas mengenai faktor-faktor klinis yang mempengaruhi kondisi pasien gagal jantung [13].

Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang menggunakan konsep entropy dan information gain untuk menentukan atribut terbaik sebagai node pemisah. Entropy digunakan untuk mengukur tingkat ketidakpastian dalam suatu dataset dan dirumuskan sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^k p_i \log_2 p_i$$

Dimana p_i adalah proporsi data kelas ke-i.

Selanjutnya, nilai information gain dihitung untuk setiap atribut:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Atribut dengan nilai information gain tertinggi dipilih sebagai node akar pada pohon keputusan. Proses ini dilakukan secara rekursif hingga seluruh data terklasifikasi atau memenuhi kondisi penghentian.

Keunggulan algoritma C4.5 antara lain mampu menangani data numerik dan kategorikal, mengelola nilai data yang hilang, serta menghasilkan model yang mudah diinterpretasikan. Oleh karena itu, algoritma ini banyak digunakan dalam penelitian klasifikasi data medis.

State of the Art

Penelitian terkait klasifikasi penyakit jantung dengan pendekatan machine learning telah banyak dilakukan dalam beberapa tahun terakhir. Berbagai algoritma seperti Naïve Bayes, Support Vector Machine, Random Forest, dan Decision Tree telah diterapkan untuk memprediksi risiko penyakit jantung dan gagal jantung dengan tingkat akurasi yang beragam [4], [8], [11]. Hasil penelitian tersebut menunjukkan bahwa algoritma decision tree memiliki keunggulan dalam hal interpretabilitas model dibandingkan algoritma lain yang bersifat black box.

Beberapa penelitian secara khusus menerapkan algoritma C4.5 dalam klasifikasi penyakit jantung dan melaporkan bahwa algoritma ini mampu menghasilkan performa yang kompetitif serta struktur pohon keputusan yang mudah dipahami oleh pengguna non-teknis, khususnya dalam konteks medis [6], [7], [13]. Selain itu, visualisasi pohon keputusan memungkinkan identifikasi atribut klinis yang paling berpengaruh dalam proses pengambilan keputusan.

Meskipun demikian, sebagian penelitian terdahulu lebih menekankan pada perbandingan algoritma tanpa melakukan analisis mendalam terhadap karakteristik dataset klinis tertentu [8], [10]. Oleh karena itu, penelitian ini berfokus pada evaluasi performa algoritma C4.5 secara spesifik pada dataset Heart Failure Clinical Records, dengan memanfaatkan RapidMiner sebagai alat bantu pemodelan dan evaluasi, sehingga diharapkan dapat memberikan kontribusi empiris yang lebih terfokus.

2Alur Penelitian

Tahapan penelitian yang dilakukan dapat diringkas sebagai berikut:

1. Pengumpulan dan pemahaman dataset Heart Failure Clinical Records.
2. Preprocessing data untuk memastikan kualitas data yang digunakan.
3. Pembagian data menjadi data latih dan data uji.
4. Pembangunan model klasifikasi menggunakan algoritma C4.5 pada RapidMiner.
5. Evaluasi performa model menggunakan confusion matrix.

HASIL DAN PEMBAHASAN

hasil penelitian berupa model klasifikasi yang dihasilkan oleh algoritma C4.5 serta evaluasi performa model berdasarkan pengujian yang telah dilakukan menggunakan RapidMiner. Selain itu, dibahas pula interpretasi hasil klasifikasi dan struktur pohon keputusan yang terbentuk. Pendekatan ini umum digunakan dalam penelitian klasifikasi data medis berbasis decision tree [6], [7], [10].

Hasil Pembentukan Model Decision Tree C4.5

Pada tahap ini dilakukan pembentukan model klasifikasi menggunakan algoritma C4.5 terhadap dataset Heart Failure Clinical Records. Proses pembentukan model dilakukan dengan menghitung nilai information gain pada setiap atribut untuk menentukan struktur pohon keputusan. Atribut yang memiliki nilai gain tertinggi dipilih sebagai simpul akar, sedangkan atribut lainnya ditempatkan pada tingkat berikutnya sesuai dengan aturan algoritma C4.5 [6], [7].

Model yang dihasilkan berupa decision tree yang merepresentasikan aturan klasifikasi berdasarkan kombinasi nilai atribut klinis pasien. Setiap cabang pada pohon keputusan menunjukkan kondisi tertentu yang mengarah pada hasil klasifikasi. Struktur ini memungkinkan model untuk mengklasifikasikan data secara sistematis dan mudah dipahami, sehingga sesuai untuk penerapan pada domain kesehatan yang membutuhkan interpretabilitas hasil [7], [14].

Visualisasi decision tree yang dihasilkan dari RapidMiner memberikan gambaran yang jelas mengenai proses pengambilan keputusan model. Dengan adanya visualisasi tersebut, hubungan antar atribut klinis dan pengaruhnya terhadap hasil klasifikasi dapat dianalisis secara lebih mendalam, sebagaimana ditunjukkan pada penelitian terdahulu yang menggunakan algoritma decision tree pada data medis [6], [13].

Hasil Pengujian Model Klasifikasi

Evaluasi performa model dilakukan menggunakan confusion matrix dan metrik akurasi. Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 60,00%. Confusion matrix yang dihasilkan ditunjukkan pada Tabel 1.

Tabel 1 Confusion Matrix Hasil Klasifikasi

	Aktual False	Aktual True
Prediksi False	51	26
Prediksi True	10	3

Berdasarkan confusion matrix tersebut, dapat diketahui bahwa:

- Sebanyak 51 data diklasifikasikan dengan benar sebagai kelas false (true negative).
- Sebanyak 3 data diklasifikasikan dengan benar sebagai kelas true (true positive).
- Terdapat 26 data yang seharusnya true namun diprediksi false (false negative).
- Terdapat 10 data yang seharusnya false namun diprediksi true (false positive).

Hasil ini menunjukkan bahwa model lebih baik dalam mengklasifikasikan kelas false dibandingkan kelas true.

Analisis Precision dan Recall

Analisis terhadap struktur decision tree yang dihasilkan menunjukkan bahwa beberapa atribut klinis muncul pada tingkat awal pohon keputusan, yang menandakan bahwa atribut tersebut memiliki kontribusi yang signifikan dalam proses klasifikasi. Hal ini sejalan dengan konsep dasar algoritma C4.5 yang memilih atribut berdasarkan nilai information gain tertinggi [6].

Visualisasi pohon keputusan memberikan kemudahan dalam menelusuri proses pengambilan keputusan model. Setiap simpul dan cabang pada pohon dapat dianalisis untuk memahami bagaimana

keputusan klasifikasi dihasilkan. Kemampuan interpretasi ini sangat penting dalam konteks aplikasi medis, karena memungkinkan hasil klasifikasi dijelaskan secara logis dan sistematis [7], [14].

Pembahasan Hasil Penelitian

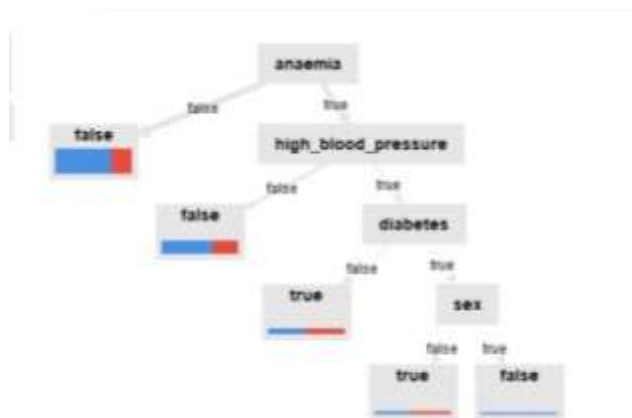
Berdasarkan hasil pengujian, algoritma C4.5 mampu menghasilkan model klasifikasi dengan tingkat akurasi sedang. Struktur pohon keputusan yang dihasilkan relatif sederhana dan mudah diinterpretasikan, sehingga menjadi keunggulan utama algoritma C4.5 dalam konteks data medis.

Namun, performa model dalam mengklasifikasikan kelas true masih rendah, terutama dilihat dari nilai recall dan precision. Hal ini menunjukkan bahwa meskipun model cukup baik dalam mengidentifikasi pasien yang tidak mengalami kejadian kematian, model belum optimal dalam mendeteksi pasien dengan risiko kematian. Kondisi ini dapat dipengaruhi oleh ketidakseimbangan jumlah data antar kelas serta keterbatasan atribut yang digunakan dalam pembentukan model.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma C4.5 dapat digunakan untuk klasifikasi gagal jantung berbasis data klinis, namun masih diperlukan pengembangan lebih lanjut, seperti penyeimbangan data atau penggunaan metode optimasi, untuk meningkatkan performa klasifikasi terutama pada kelas minoritas.

Visualisasi Decision Tree Algoritma C4.5

Hasil pemodelan menggunakan algoritma C4.5 menghasilkan sebuah struktur pohon keputusan (decision tree) yang menggambarkan proses pengambilan keputusan dalam klasifikasi pasien gagal jantung. Visualisasi pohon keputusan hasil pemodelan ditunjukkan pada Gambar 1.



Gambar 1 Visualisasi Pohon Keputusan Algoritma C4.5 pada Dataset Heart Failure Clinical Records

Pada Gambar 1, terlihat bahwa atribut anaemia menjadi node akar pada pohon keputusan, yang menunjukkan bahwa atribut tersebut memiliki nilai information gain tertinggi. Selanjutnya, atribut high_blood_pressure, diabetes, dan sex digunakan sebagai node pemisah pada level berikutnya. Struktur pohon keputusan tersebut menunjukkan bahwa kombinasi beberapa atribut klinis berperan dalam menentukan hasil klasifikasi pasien.

Visualisasi decision tree memberikan kemudahan dalam memahami alur klasifikasi yang dihasilkan oleh algoritma C4.5. Setiap percabangan pada pohon keputusan merepresentasikan kondisi klinis tertentu yang mengarah pada kelas akhir. Dengan demikian, model yang dihasilkan tidak hanya memberikan hasil prediksi, tetapi juga menyediakan informasi yang bersifat interpretatif bagi pengguna.

Diskusi Hasil Penelitian

Hasil pengujian menunjukkan bahwa algoritma C4.5 mampu membangun model klasifikasi gagal jantung berbasis data klinis dengan tingkat akurasi yang tergolong sedang. Model yang dihasilkan cenderung lebih efektif dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas, sebagaimana tercermin dari nilai recall yang lebih tinggi pada kelas tersebut [8], [10].

Visualisasi decision tree memberikan gambaran yang jelas mengenai hubungan antar atribut klinis dalam proses pengambilan keputusan. Atribut seperti anaemia, high blood pressure, diabetes, dan sex muncul sebagai faktor penting dalam penentuan kelas. Hal ini menunjukkan bahwa algoritma C4.5 mampu mengekstraksi pola yang relevan dari data klinis, meskipun performanya masih dipengaruhi oleh distribusi data yang tidak seimbang [6], [7].

Jika dibandingkan dengan penelitian terdahulu yang menerapkan algoritma C4.5 pada kasus klasifikasi penyakit jantung, hasil penelitian ini menunjukkan kecenderungan yang sejalan, khususnya dalam hal kemudahan interpretasi model. Perbedaan nilai performa yang diperoleh mengindikasikan bahwa karakteristik dataset dan tahapan pengolahan data memiliki pengaruh signifikan terhadap hasil klasifikasi [6], [10].

Implikasi Hasil Penelitian

Implikasi dari hasil penelitian ini menunjukkan bahwa algoritma C4.5 berpotensi digunakan sebagai alat bantu awal dalam analisis klasifikasi gagal jantung berbasis data klinis. Model pohon keputusan yang dihasilkan dapat membantu dalam memahami faktor-faktor klinis yang berkontribusi terhadap kondisi pasien, sehingga dapat memberikan wawasan tambahan bagi proses analisis data kesehatan.

Namun demikian, keterbatasan performa model dalam mengklasifikasikan kelas minoritas menunjukkan bahwa penerapan algoritma C4.5 secara langsung belum sepenuhnya optimal untuk tujuan prediksi risiko secara presisi. Oleh karena itu, hasil penelitian ini mengindikasikan perlunya pendekatan lanjutan, seperti peningkatan kualitas data, penyeimbangan kelas, atau integrasi dengan metode klasifikasi lain. Implikasi ini membuka peluang penelitian selanjutnya untuk mengembangkan model klasifikasi yang lebih akurat dan andal dalam mendukung pengambilan keputusan di bidang kesehatan.

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma C4.5 pada klasifikasi data gagal jantung berbasis Heart Failure Clinical Records mampu menghasilkan model klasifikasi yang bersifat interpretatif melalui struktur pohon keputusan. Hasil pengujian menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan pasien pada kelas mayoritas, yang tercermin dari nilai akurasi dan recall yang relatif lebih tinggi pada kelas tersebut. Selain itu, visualisasi decision tree memperlihatkan bahwa atribut klinis seperti anaemia, high blood pressure, diabetes, dan sex memiliki pengaruh dalam

proses pengambilan keputusan model, sebagaimana juga ditemukan pada penelitian terdahulu di bidang klasifikasi penyakit jantung [6], [7].

Meskipun demikian, performa model dalam mengklasifikasikan kelas minoritas masih tergolong rendah, yang menunjukkan adanya keterbatasan dalam mendeteksi pasien dengan risiko yang lebih tinggi. Kondisi ini mengindikasikan perlunya pengembangan lebih lanjut, baik dari sisi pengolahan data maupun metode klasifikasi, agar model yang dihasilkan memiliki performa yang lebih seimbang. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma C4.5 dapat digunakan sebagai pendekatan awal dalam analisis klasifikasi gagal jantung berbasis data klinis, sekaligus memberikan dasar bagi penelitian lanjutan untuk meningkatkan akurasi dan keandalan model klasifikasi di bidang kesehatan [8], [10].

SARAN

Penelitian selanjutnya disarankan untuk melakukan pengolahan data yang lebih mendalam, khususnya dalam menangani ketidakseimbangan kelas pada dataset gagal jantung, agar performa model dalam mengklasifikasikan kelas minoritas dapat ditingkatkan. Selain itu, penggunaan metode seleksi atribut atau teknik feature engineering dapat dipertimbangkan untuk mengoptimalkan atribut yang paling berpengaruh terhadap hasil klasifikasi. Penelitian lanjutan juga dapat membandingkan algoritma C4.5 dengan algoritma klasifikasi lainnya, seperti Random Forest, Support Vector Machine, atau Naïve Bayes, guna memperoleh model dengan performa yang lebih optimal. Penggunaan teknik validasi yang lebih bervariasi serta pemanfaatan metrik evaluasi tambahan juga diharapkan dapat memberikan gambaran performa model yang lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan dan masukan yang berharga sehingga penelitian ini dapat diselesaikan dengan baik. Selain itu, penulis mengapresiasi pihak penyedia dataset Heart Failure Clinical Records serta pengembang perangkat lunak RapidMiner yang telah mendukung proses pengolahan dan analisis data dalam penelitian ini.

DAFTAR PUSTAKA

- [1] A. Chicco and G. Jurman, "Machine Learning Can Predict Survival of Patients with Heart Failure," IEEE Access, vol. 8, pp. 1–12, 2020. Link: <https://arxiv.org/pdf/1902.02276.pdf>
- [2] D. Dua and C. Graff, "Heart Failure Clinical Records Dataset," UCI Machine Learning Repository, Irvine, CA, USA, 2020. Link: <https://archive.ics.uci.edu/ml/datasets/heart+failure+clinical+records>
- [3] Y. Sun, L. Wang, and M. Li, "A Deep Learning Method for Heart Failure Survival Prediction," Journal of Healthcare Engineering, vol. 2021, Article ID 6696989, 2021. Link: <https://downloads.hindawi.com/journals/jhe/2021/6696989.pdf>

- [4] A. M. Abdulazeez and S. S. Hasan, "Classification of Heart Diseases Based on Machine Learning: A Review," *International Journal of Informatics and Information System Computer Engineering*, vol. 6, no. 1, pp. 31–52, Aug. 2025. Link: <https://ojs.unikom.ac.id/index.php/injiiscom/article/view/13600>
- [5] A. R. Ilyas, S. Javaid, and I. L. Kharisma, "Heart Disease Prediction Using Machine Learning," *Engineering Proceedings*, vol. 107, no. 1, Article 124, 2025. Link: <https://www.mdpi.com/2673-4591/107/1/124/pdf>
- [6] B. S. Reddy and P. R. Samuel, "Classification of Heart Disease Using C4.5 and CART Decision Tree Algorithms," *Journal of Statistics and Management Systems*, vol. 24, no. 4, pp. 741–755, 2021. Link: https://www.researchgate.net/publication/353858235_Classification_of_Heart_Disease_using_C45_and_CART_Decision_Tree_Algorithms.pdf
- [7] S. K. Polat and H. D. Ozkan, "Heart Disease Classification Using Decision Tree-Based Algorithms," *Diagnostics*, vol. 11, no. 11, Article 2074, 2021. : <https://www.mdpi.com/2075-4418/11/11/2074/pdf>
- [8] C. A. Singh et al., "Comparative Analysis of Machine Learning Techniques for Heart Failure Prediction," *Applied Sciences*, vol. 12, no. 7, Article 3637, 2022. Link: <https://www.mdpi.com/2076-3417/12/7/3637/pdf>
- [9] S. K. Dwivedi and H. S. Rautaray, "Feature Selection Framework for Heart Disease Classification Using Hybrid Machine Learning," *International Journal of Computer Science and Engineering*, vol. 9, no. 8, pp. 102–110, 2021. Link: https://www.researchgate.net/publication/353423798_Feature_Selection_Framework_for_Heart_Disease_Classification_Using_Hybrid_Machine_Learning
- [10] A. Tavana et al., "Heart Failure Clinical Data Classification Using Advanced Machine Learning Techniques," *Scientific Reports*, vol. 13, Article 35806, 2023. Link: <https://www.nature.com/articles/s41598-023-35806-6.pdf>
- [11] M. A. Alotaibi et al., "Machine Learning Models for Heart Disease Prediction: A Comparative Study," *Healthcare*, vol. 10, no. 11, Article 2222, 2022. Link: <https://www.mdpi.com/2227-9032/10/11/2222/pdf>
- [12] P. Gupta and A. R. Sharma, "Performance Evaluation of Machine Learning Techniques for Heart Disease Detection," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 2751–2760, 2021. Link: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9272023>
- [13] S. M. R. Islam, M. M. Hasan, and M. R. Karim, "A Decision Tree-Based Approach for Heart Disease Classification," *Diagnostics*, vol. 12, no. 9, Article 2142, 2022. Link: <https://www.mdpi.com/2075-4418/12/9/2142/pdf>
- [14] A. A. Khan et al., "Optimized Decision Tree for Healthcare Classification Problems," *IEEE Access*, vol. 9, pp. 1–15, 2021. Link: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9497179>
- [15] D. Islamuddin et al., "Machine Learning-Based Heart Disease Detection Using Open Source Tools," *Sensors*, vol. 22, no. 5, Article 1836, 2022. Link: <https://www.mdpi.com/1424-8220/22/5/1836/pdf>