Terbit online pada laman web jurnal: **https://jurnal.plb.ac.id/index.php/tematik/index**

T E M A T I K

Jurnal **Teknologi Informasi Komunikasi (e-Journal)**

Vol. 12 No. 2 (2025) 164 - 172 ISSN Media Elektronik: 2443-3640

Analisis Algoritma Machine Learning untuk Gagal Jantung dengan *Interpretability* SHAP dan LIME

I Gede Sugita Aryandana[1], Leni Anggraini Susanti[2], Putu Eka Suryadana[3], Wayan Gede Suka Parwita[4]
[1,2,3]Administrasi Bisnis, Bisnis Digital, Politeknik Negeri Bali
[4]Administrasi Bisnis, Manajemen Bisnis International, Politeknik Negeri Bali
[1]sugitaaryandana@pnb.ac.id, [2]lenissnti@pnb.ac.id, [3]esuryadana@pnb.ac.id, [4]gede.suka@pnb.ac.id

Abstract

Health is a state of complete physical, mental, and social well-being that enables individuals to lead productive lives. Among the various dimensions of health, physical health is the most fundamental as it plays a vital role in maintaining the quality of human life. One of the primary indicators of physical health is heart function, which is crucial in supporting bodily activities and overall performance. Impairments in heart function not only affect physical conditions but can also negatively impact mental and social well-being, ultimately leading to heart failure. Early detection of heart failure is essential, prompting the need for research on the application of machine learning algorithms as analytical methods to enhance accuracy and efficiency in the diagnostic process. In this study, SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were utilized as interpretability tools, employing medical datasets from heart failure patients as the analysis objects. The findings indicate that the random forest algorithm is the most effective in predicting heart failure risk, yielding an overall accuracy of 0.8334, an average precision of 0.8125, a recall of 0.793, an F1-score of 0.8015, and an area under the curve (AUC) value of 0.90. Based on the classification results, interpretability analyses were conducted using SHAP and LIME. The SHAP method provided comprehensive insights into the influence of key variables such as time, serum creatinine, ejection fraction, and age on the risk of heart failure. Meanwhile, LIME offered localized explanations, highlighting serum creatinine as the main variable increasing the prediction of mortality in specific cases. These interpretability methods contribute to greater transparency and trust in machine learning-based medical diagnoses.

Keywords: Heart Failure, Machine Learning, SHAP, LIME

Abstrak

Kesehatan adalah keadaan sejahtera secara fisik, mental, dan sosial yang memungkinkan setiap individu dapat menjalani kehidupan secara produktif. Di antara berbagai dimensi kesehatan, kesehatan fisik menjadi aspek yang paling fundamental karena berperan penting dalam menjaga kualitas hidup manusia. Salah satu indikator penting dalam kesehatan fisik adalah fungsi jantung, sehingga deteksi dini terhadap risiko gagal jantung menjadi krusial untuk mencegah kondisi klinis yang lebih berat. Penelitian ini bertujuan mengembangkan model prediksi gagal jantung menggunakan pendekatan *Machine Learning* dengan membandingkan tiga algoritma, yaitu *Random Forest, Logistic Regression*, *dan Decision Tree*. *Dataset* medis pasien penderita gagal jantung digunakan sebagai objek analisis untuk mengevaluasi performa model berdasarkan metrik akurasi, *precision*, *recall*, *f1-score*, dan *AUC*. Hasil penelitian menunjukkan bahwa *Random Forest* merupakan algoritma dengan performa terbaik dengan nilai akurasi 0,8334, *precision* 0,8125, *recall* 0,793, f1-score 0,8015, dan AUC 0,90. Selanjutnya, interpretabilitas model terbaik dianalisis menggunakan metode *SHAP* dan *LIME* untuk menjelaskan kontribusi variabel terhadap prediksi. Metode *SHAP* mengidentifikasi *variabel time*, *serum creatinine*, *ejection fraction*, dan *age* sebagai faktor paling berpengaruh secara global, sedangkan *LIME* memberikan penjelasan lokal yang konsisten, terutama menyoroti serum *creatinine* sebagai indikator kuat peningkat risiko kematian. Penelitian ini menunjukkan bahwa model prediktif yang akurat, dipadukan dengan metode interpretabilitas, mampu memberikan pemahaman yang transparan mengenai faktor risiko gagal jantung. Temuan ini dapat mendukung tenaga medis dalam pengambilan keputusan klinis berbasis data dan meningkatkan ketepatan diagnosis.

Kata kunci: Gagal Jantung, *Machine Learning, SHAP, LIME*

1. Pendahuluan

Kesehatan merupakan kondisi holistik yang mencakup kesejahteraan fisik, mental, dan sosial, sehingga memungkinkan individu untuk menjalani kehidupan yang bermakna dan produktif [1], [2], [3] . Dari ketiga dimensi tersebut, kesehatan fisik menjadi aspek yang paling mendasar karena secara langsung memengaruhi kapasitas tubuh dalam menjalankan aktivitas sehari-hari [4]. Kesehatan fisik yang baik tidak hanya berfungsi

Diterima Redaksi: 29-09-2025 | Selesai Revisi: 28-11-2025 | Diterbitkan Online: 19-12-2025

sebagai penopang produktivitas, tetapi juga menjadi indikator utama kualitas hidup masyarakat [5].

Salah satu tantangan terbesar dalam menjaga kesehatan fisik masyarakat adalah meningkatnya prevalensi penyakit *kardiovaskular*, khususnya gagal jantung. Gagal jantung merupakan salah satu penyebab utama kematian di dunia, termasuk di Indonesia. Organisasi Kesehatan Dunia (WHO) mencatat bahwa penyakit jantung koroner dan gagal jantung menyumbang angka mortalitas yang tinggi setiap tahunnya [6], [7].

Kondisi ini diperparah oleh faktor risiko seperti hipertensi, diabetes melitus, obesitas, kebiasaan merokok, serta pola hidup sedentari yang kian meluas akibat modernisasi [8], [9]. Selain berdampak pada angka kematian, gagal jantung juga berdampak pada kondisi fisik yang mengganggu kesejahteraan mental dan sosial seseorang, yang pada akhirnya memperburuk kondisi jantung. Oleh karena itu, deteksi dini menjadi aspek penting dalam upaya menekan angka kejadian dan meningkatkan kualitas hidup pasien dengan bantuan teknologi [10], [11].

Perkembangan teknologi kesehatan, khususnya pemanfaatan data medis, memberikan peluang baru dalam mendukung deteksi dan analisis gagal jantung [12]. Data medis pasien dapat dianalisis secara komprehensif untuk memprediksi risiko, mengidentifikasi pola gagal jantung, serta memberikan rekomendasi medis yang lebih tepat [13]. Pada konteks ini, algoritma *Machine Learning* seperti *Random Forest, Logistic Regression*, dan *Decision Tree* banyak digunakan karena kemampuannya mengolah variabel klinis untuk menghasilkan prediksi risiko secara lebih efisien. Namun, meskipun model-model ini menunjukkan performa yang kuat, aspek transparansi (*interpretability*) masih menjadi tantangan, terutama pada algoritma kompleks seperti *random forest* yang bersifat *black box*. [14].

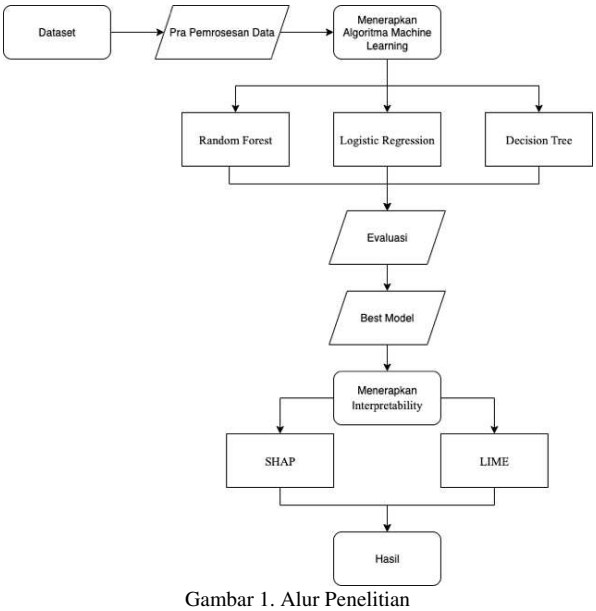
Penelitian-penelitian sebelumnya menegaskan bahwa model ensemble seperti *Random Forest* dan *XGBoost* mampu memberikan prediksi akurat, tetapi sulit diinterpretasikan secara klinis [15], [16]. Sebaliknya, *Logistic Regression* lebih mudah dipahami namun kadang kalah dalam akurasi [17]. Kesenjangan ini menunjukkan perlunya studi komparatif yang tidak hanya menilai kinerja model, tetapi juga membandingkan tingkat interpretabilitasnya sehingga model yang dipilih tidak hanya akurat tetapi juga dapat dipahami oleh tenaga medis [18].

Untuk menjembatani masalah tersebut, dibutuhkan kajian mengenai *interpretability* yang tidak hanya mengevaluasi hasil prediksi, tetapi juga membantu menentukan pendekatan terbaik dalam memprediksi gagal jantung. [19].

Penelitian ini mengintegrasikan dua metode interpretabilitas, yaitu *SHAP* dan *LIME*. *SHAP* digunakan untuk melihat kontribusi variabel secara global berdasarkan nilai *Shapley*, sedangkan *LIME* memberikan interpretasi lokal pada setiap prediksi kasus individu. Kombinasi keduanya memungkinkan pemahaman yang lebih komprehensif, baik pada level keseluruhan model maupun kasus-per-kasus, sehingga mendukung transparansi klinis yang dibutuhkan [20]. Berdasarkan hal tersebut, penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma machine learning yaitu Random Forest, Logistic Regression, dan *Decision Tree* dalam memprediksi risiko gagal jantung, serta mengevaluasi interpretabilitas model terbaik menggunakan *SHAP* dan *LIME*. [21]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam penyediaan pendekatan prediksi yang tidak hanya akurat tetapi juga transparan dan dapat diterapkan dalam pengambilan keputusan klinis.

2. Metode Penelitian

Penelitian ini bertujuan untuk menentukan faktor utama yang berkontribusi terhadap resiko gagal jantung dengan menggunakan algoritma *machine learning* seperti *random forest*, *logistic regression*, dan *decision tree* serta menggunakan *interpretability* untuk mengukur *feature importance* [22][23]. Tahapan penelitian yang jelas dan terstruktur sangat penting untuk memastikan hasil prediksi menjadi transparan dan mudah dipahami. Tahapan penelitian yang diterapkan kemudian divisualisasikan pada Gambar 1.



Gambar 1. Alur Penelitian

Diagram pada Gambar 1 menggambarkan alur kerja penelitian dalam membangun model prediksi gagal jantung menggunakan tiga algoritma *Machine Learning* serta interpretabilitas model. Alur prosesnya adalah sebagai berikut:

DOI: https://doi.org/10.38204/tematik

Dataset yang digunakan dalam penelitian ini merupakan data medis dari *Kaggle* pada tahun 2024 yang telah difinalisasi dari berbagai sumber, seperti data medis dari *Cleveland*, *Hungarian* dan *Switzerland*. Data ini digunakan untuk menguji keakuratan proses prediksi algoritma *machine learning*, serta mengevaluasi hasil tersebut dengan menggunakan *interpretability* untuk mendapatkan hasil prediksi yang lebih transparan.

Pra pemrosesan data dalam penelitian ini merupakan tahapan penting yang mencakup proses, *standart scaller* dan pembagian data menjadi *train testing* dari data medis. Variabel yang ada di dalam dataset kemudian diolah agar mampu merepresentasikan karakteristik data yang relevan, sehingga hasil prediksi algoritma *machine learning* dapat lebih akurat dan sesuai dengan tujuan penelitian.

Tahap selanjutnya adalah proses menerapkan algoritma *machine learning* menggunakan algoritma *random forest*, *logistic regression*, dan *decision tree*.

Algoritma *random forest* digunakan dalam penelitian ini, dengan tujuan meningkatkan efektivitas prediksi serta mendukung proses pengambilan keputusan medis berbasis data.

Algoritma *logistic regression* digunakan untuk menganalisis faktor medis yang berkontribusi terhadap gagal jantung, dengan tujuan membangun model prediksi yang mampu mengklasifikasikan pasien berisiko tinggi dan rendah berdasarkan variabel dari dataset yang telah disediakan.

Algoritma *decision tree* digunakan dalam penelitian ini untuk mengklasifikasikan kondisi pasien berdasarkan variabel medis, seperti *age*, *diabetes*, *sex,* serta *smoking*, sehingga dapat membantu memperkirakan risiko seseorang mengalami gagal jantung.

Tahap evaluasi dalam penelitian ini, dilakukan penilaian kinerja berbagai model *machine learning* dengan menggunakan serangkaian metrik evaluasi yang komprehensif, meliputi *accuration*, *precision*, *recall*, *F1-score*, serta kurva *AOC* (*Area Under Curve*) - ROC (*Receiver Operating Characteristic*). Metrik ini dipilih karena mampu memberikan gambaran menyeluruh mengenai kemampuan algoritma dalam melakukan prediksi. Proses evaluasi ini tidak hanya membandingkan performa setiap algoritma secara kuantitatif, tetapi juga memastikan bahwa algoritma yang dipilih memiliki tingkat konsistensi tinggi dan minim risiko kesalahan prediksi. Dengan pendekatan ini, algoritma *machine learning* memiliki hasil prediksi terbaik. Untuk melihat hasil prediksi dibutuhkan *confusion matrix* dalam mengukur dan mengevaluasi kinerja klasifikasi. *Confusion matrix* berguna untuk membandingkan hasil prediksi secara aktual dalam dataset. Komponen pada Tabel 1 digunakan untuk perbandingan yang melibatkan jumlah nilai pada TP (*True Positive*), FP(*False Positive*), TN(*True Positive*), dan FN (*False Negative*) [24].

Tabel 1. *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Proses pemilihan *best* model *machine learning* dilakukan melalui serangkaian tahapan evaluasi yang ketat dengan memanfaatkan berbagai metrik performa seperti metrik *accuration*, *precision*, *recall*, serta *F1-score*, dan *AUC - ROC*.

Metrik *Accuration* digunakan dalam penelitian ini untuk memprediksi risiko gagal jantung. Melalui penelitian ini, tingkat akurasi yang dihasilkan dalam mengklasifikasikan pasien berisiko dapat terlihat dengan jelas, sehingga memberikan gambaran umum terkait prediksi yang dihasilkan. Adapun rumus yang digunakan untuk mendapatkan hasil metrik *accuration*. Variabel (TP dan TN) pada Rumus 1 berguna untuk menilai perbandingan antara porsi prediksi yang akurat dengan total yang dihasilkan dengan variabel (TP, TN. FP, FN) [24].

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Metrik *Precision* digunakan dalam penelitian ini untuk memprediksi risiko gagal jantung. Metrik *precision* berfungsi untuk mengukur seberapa akurat metrik dalam mengklasifikasikan pasien yang benar-benar berisiko dari seluruh pasien yang diprediksi berisiko. Dengan demikian, hasil ini memberikan gambaran mengenai kemampuan metrik dalam meminimalkan *false positive* sehingga prediksi yang dihasilkan lebih relevan dan terpercaya. Adapun rumus yang digunakan untuk mendapatkan hasil metrik *precision*. Variabel (TP) pada Rumus 2 berguna untuk menilai seberapa besar porsi prediksi data dengan nilai positip dengan nilai yang sebenarnya yang kemudian dibandingkan dengan keseluruhan hasil prediksi yang bersifat positip (TP dan FP) [24].

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Metrik *Recall* digunakan dalam penelitian ini untuk memprediksi risiko gagal jantung. *Recall* berfungsi untuk mengukur sejauh mana metrik mampu menemukan seluruh pasien yang termasuk kategori berisiko gagal jantung dari keseluruhan pasien berisiko yang ada. Dengan demikian, metrik ini menekankan pada kemampuan dalam meminimalkan kesalahan *false negative*, sehingga pasien berisiko tidak terlewat dalam proses klasifikasi. Adapun rumus yang digunakan untuk mendapatkan hasil metrik *recall*. Variabel (TP) pada Rumus 3 berguna untuk mengukur porsi data yang diprediksi sebagai positip yang kemudian

DOI: https://doi.org/10.38204/tematik

dikombinasikan dengan total data yang sebenarnya bernilai positip (TP dan FN) [24].

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Metrik *F1-Score* digunakan dalam penelitian ini untuk memprediksi risiko gagal jantung. *F1-Score* berfungsi sebagai tolok ukur keseimbangan antara *precision* dan *recall*, dengan cara menghitung rata-rata dari keduanya. Metrik ini penting karena memberikan gambaran menyeluruh tentang performa model, khususnya ketika distribusi data pasien berisiko dan tidak berisiko tidak seimbang, sehingga hasil prediksi tetap adil dan representatif. Adapun Rumus 4 digunakan untuk mendapatkan hasil metrik *F1-Score*. Rumus metrik *F1-Score* digunakan untuk mengukur rata-rata perbandingan antara nilai *precision* dan *recall* [24].

$$F1 - Score = 2\,x\,\frac{PrecisionxRecall}{Precision+Recall} \quad (4)$$

Metrik *Area Under Curve - Receiver Operating Characteristic* (AUC - *ROC*) digunakan untuk menilai efektivitas model klasifikasi bersifat efektif atau tidak efektif [24].

Dari hasil evaluasi tersebut, dipilihlah model terbaik yang mampu menunjukkan kemampuan paling unggul dalam memprediksi data dengan konsistensi tinggi. Dengan demikian, model ini diharapkan dapat memberikan solusi yang efektif dan andal untuk mengatasi permasalahan yang dikaji dalam penelitian ini.

Penelitian ini menerapkan *interpretability* menggunakan metode *SHAP* dan *LIME* untuk meningkatkan transparansi hasil prediksi *machine learning*.

Model *SHAP* digunakan untuk memberikan penjelasan yang mendalam dan komprehensif terhadap prediksi, khususnya dalam studi kasus gagal jantung. SHAP mampu mengukur kontribusi setiap variabel secara global, sehingga membantu memahami faktor risiko yang memengaruhi terjadinya gagal jantung. Hal ini menjadikan proses *interpretability* tidak hanya menghasilkan prediksi yang akurat, tetapi juga menyediakan penjelasan yang jelas.

Sementara itu, model *LIME* berfungsi untuk mengidentifikasi faktor utama yang memengaruhi prediksi risiko gagal jantung pada tingkat individual pasien dengan memperhatikan variabel penting seperti tekanan darah, kadar kolesterol, dan usia. Pendekatan ini memungkinkan tenaga medis memahami keputusan dari model yang bersifat black-box secara lebih transparan, meningkatkan kepercayaan, serta mendukung pengambilan keputusan klinis yang lebih tepat dan personal.

Pada tahap hasil, penelitian memasuki proses akhir. Hasil pengujian menggunakan *interpretability* metode *SHAP* dan LIME yang telah dilatih kemudian dianalisis dan disajikan sebagai laporan penelitian. Studi ini ditutup dengan penyebarluasan hasil analisis menggunakan metode *SHAP* dan *LIME* sebagai *interpretability*, sehingga temuan yang diperoleh dapat memperkaya pengetahuan sekaligus mendorong penerapan *machine learning* di masa mendatang. Pada fase akhir ini, disusun kerangka sistematis yang memastikan hasil prediksi bersifat valid dan transparan, sehingga dapat dijadikan acuan dalam pengembangan model maupun pengambilan keputusan berbasis data yang lebih akurat.

3. Hasil dan Pembahasan

Rangkaian proses penelitian dilakukan secara bertahap untuk membangun analisa model yang mampu mengidentifikasi variabel penyebab gagal jantung berdasarkan dataset yang telah diuji. Dimulai dari proses pra-pemrosesan data sampai hasil training algoritma *machine learning* dengan *interpretability* metode *SHAP* dan metode *LIME*. Hasil dari setiap tahapan penelitian menunjukkan proses transformasi data serta pemilihan algoritma yang berperan penting dalam menentukan interpretability untuk memperoleh prediksi yang akurat dan transparan.

3.1 Pra Pemrosesan Data

Tahap selanjutnya dalam pra-pemrosesan data adalah dengan membagi dataset menjadi dua bagian yaitu 80% data *training* dan 20% data *testing* dengan menerapkan *cross validation* (*cv*) 5. Pada tahap pra-pemrosesan data dilakukan normalisasi untuk melengkapi variabel serta menghapus nilai yang dianggap tidak relevan. Tabel 1 menampilkan variabel sebelum normalisasi, sedangkan Tabel 2 memperlihatkan hasil setelah proses normalisasi.

Pada Tabel 2, terdapat 13 variabel yang menjadi bagian dari dataset. Variabel tersebut terdiri dari *age, anaemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum sodium, sex, smoking, time* dan *death event*. Selanjutnya, variabel diubah dan disederhanakan menggunakan *standard scaler* dengan tujuan menyetarakan skala antarvariabel. Proses ini dilakukan dengan mentransformasi setiap variabel agar memiliki rata-rata (*mean*) 0 dan standar deviasi 1 melalui perhitungan *z-score*, yaitu selisih nilai data terhadap rata-rata yang dibagi dengan standar deviasi. Dengan transformasi data tersebut, variabel yang semula memiliki satuan dan rentang berbeda dapat disejajarkan dalam skala yang sama, sehingga tidak ada variabel yang mendominasi analisis akibat perbedaan skala pengukuran. Deskripsi dan aturan transformasi data yang dijalankan tercantum pada Tabel 2.

Transformasi data yang tercantum pada Tabel 2 mengindikasikan 13 variabel yang dijadikan dasar

DOI: https://doi.org/10.38204/tematik

dalam menganalisis sejumlah variabel. Masing-masing variabel diubah ke bentuk nilai numerik diskrit agar dapat merepresentasikan kategori tertentu. Dengan cara ini, hubungan antar variabel dapat divisualisasikan secara lebih sederhana, sekaligus memungkinkan analisis data berlangsung lebih efisien dan mendalam. Pada Tabel 2 variabel *death event* dihilangkan karena variabel tersebut merupakan variabel y.

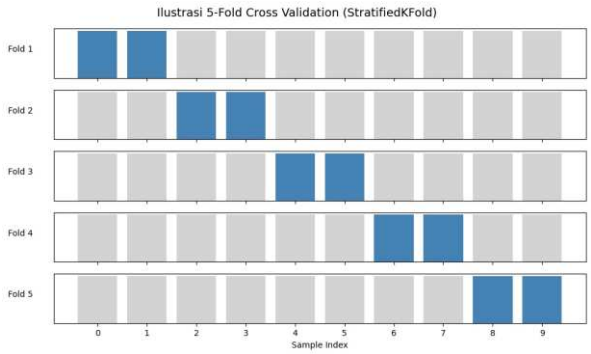
Tabel 2. Hasil Perolehan Data

Age	Anaemia	creatinine_phosphokinase	Diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	Sex	Smoking	Time	Death_Event
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358.0	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	1

Tabel 3. Data Setelah di Normalisasi

Age	Anaemia	creatinine_phosphokinase	Diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	Sex	Smoking	Time
1.17	-0.87	0.6916146	-847.579	-1.773	1.3592	0.110527	1.21222	-1.468519	0.7	-0.6876	-2.04
-0.4	-0.87	2.401.701	-847.579	0.100	-0.7356	0.093441	-0.0876	-0.244180	0.7	-0.6876	-1.97
0.43	-0.87	-0.5534.2	-847.579	-1.773	-0.7356	-1.09314	0.38181	-1.642142	0.7	1.45416	-1.94
-0.9	1.14	-0.8338.8	-847.579	-1.773	-0.7356	-0.49471	1.21222	-0.006502	0.7	-0.68768	-1.94
0.43	1.14	-4.623.35	1.179830	-1.773	-0.7356	0.7202.7	1.71506	-3.285072	-1	-0.68768	-1.90

Tahap berikutnya adalah proses *cross-validation* yang dilakukan setelah dataset melalui tahap normalisasi. Adapun ilustrasi *cross validation* yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. *Cross Validation (CV)* 5

Pada Gambar 2, keseluruhan dataset dibagi menjadi lima bagian (*fold*) dengan ukuran yang relatif sama untuk dilakukan data *training* dan data *testing*. Proses *data training* dan data *testing* dilakukan secara bergantian sebanyak lima kali, yang dimana pada setiap iterasi satu *fold* digunakan sebagai data *testing*, sedangkan empat fold lainnya digunakan sebagai data *training*. Dengan demikian, seluruh data memiliki kesempatan yang sama untuk digunakan sebagai data *testing*. Hasil evaluasi model diperoleh dari rata-rata skor performa pada kelima iterasi tersebut, sehingga dapat mengurangi potensi bias akibat pembagian data tertentu dan memberikan estimasi yang lebih stabil terhadap kemampuan generalisasi model.

3.2 Menerapkan Algoritma *Machine Learning*

Tahap setelah pra pemrosesan data adalah menerapkan algoritma *machine learning* untuk membangun model prediktif yang mampu melakukan klasifikasi sesuai dengan tujuan penelitian. Algoritma yang digunakan dalam pembentukan model ini mencakup algoritma *random forest*, *logistic regression*, dan *decision tree*

Algoritma *random forest* digunakan dalam penelitian ini untuk proses *training dataset* dengan tujuan memperoleh dan membangun model klasifikasi yang akurat, sehingga menghasilkan prediksi yang baik. Adapun data *training* yang sudah diterapkan dengan menggunakan algoirtma *random forest* dapat dilihat pada Tabel 4.

Tabel 4. *Classification Data Training Random Forest*

Class	*Accuracy*	*Precission*	*Recall*	*F1-score*
0		0.974	0.914	0.943
1	0.924686	0.839	0.948	0.890

Pada Tabel 4, hasil data *training* menunjukkan bahwa untuk *class* 0 dan *class 1* didapatkan nilai *accuracy* sebesar 0,9246. Nilai tersebut mengindikasikan bahwa model memiliki tingkat *accuracy* dan performa yang tinggi, baik dalam mengenali maupun memprediksi kategori yang ada.

DOI: https://doi.org/10.38204/tematik

Algoritma logistic regression digunakan dalam penelitian ini untuk untuk menganalisis berbagai faktor medis yang berperan dalam menyebabkan gagal jantung, dengan maksud membangun hasil model prediksi yang akurat dengan mengklasifikasikan pasien ke dalam kelompok dengan risiko yang lebih tinggi maupun rendah berdasarkan variabel yang ada di dalam dataset. Adapun data *training* yang sudah diterapkan dengan menggunakan algoirtma *logistic regression* dapat dilihat pada Tabel 5

Tabel 5. *Classification Data Training Logistic Regression*

Class	*Accuracy*	*Precission*	*Recall*	*F1-score*
0		0.923	0.815	0.866
1	0.828452	0.688	0.857	0.763

Pada Tabel 5, Hasil yang didapat mengindikasikan bahwa model memiliki kemampuan yang cukup baik dalam mengenali *class* 0 dengan nilai precision dan F1-score yang tinggi, namun performa pada *class* 1 cenderung lebih rendah, terutama pada nilai *precision*. Perbedaan ini dapat menggambarkan adanya ketidakseimbangan dalam kemampuan model memprediksi kedua kelas pada data yang di *training*.

Algoritma *decision tree* digunakan dalam penelitian ini untuk mengklasifikasikan kondisi pasien berdasarkan variabel medis seperti *age, sex, smoking, serta* diabetes. Dengan pendekatan ini, diharapkan dapat diperoleh model prediksi yang efektif dalam memperkirakan risiko gagal jantung pada pasien. Adapun data *training* yang sudah diterapkan dengan menggunakan algoritma *decision tree* dapat dilihat pada Tabel 6

Tabel 6. *Classification Data Training Decision Tree*

Class	*Accuracy*	*Precission*	*Recall*	*F1-score*
0		0.929	0.895	0.912
1	0.8828845	0.795	0.857	0.825

Pada Tabel 6, hasil evaluasi model menunjukkan bahwa untuk *class* 0 dan *class* 1 memiliki, nilai *accuration* sebesar 0,8828845. Hasil ini memperlihatkan bahwa *logistic regression* memiliki kemampuan yang baik dalam mengklasifikasikan kedua *class*, dengan performa yang lebih tinggi di *class* 0 dibandingkan *class* 1, namun secara keseluruhan model cukup efektif dalam mengidentifikasi pasien berisiko berdasarkan variabel dataset yang telah dianalisis.

3.3. Evaluasi

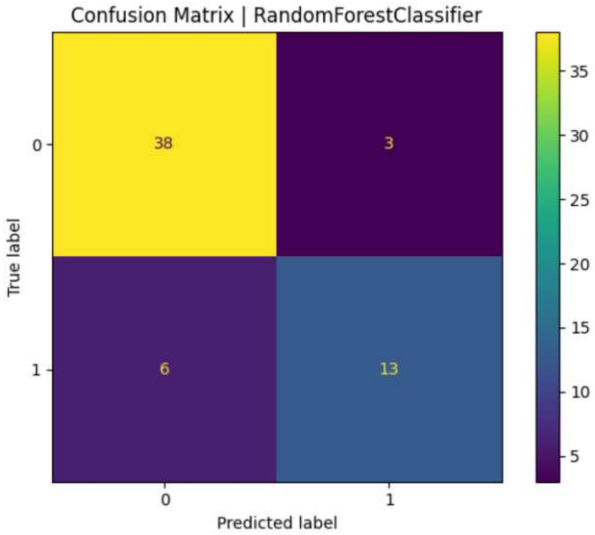
Tahap selanjutnya dalam penelitian ini adalah melakukan evaluasi berdasarkan hasil pengujian menggunakan data *training*. Mengingat hasil yang diperoleh pada data *training* menunjukkan performa yang cukup memuaskan untuk algoritma *random forest*, maka pengujian dilanjutkan dengan menggunakan data *testing* dan algoritma *random forest*. Tujuan dilakukan pengujian lanjutan adalah untuk mendapatkan hasil yang lebih baik dan membandingkan kemampuan klasifikasi masing-masing model secara objektif. Adapun hasil pengujian menggunakan data *testing* dapat dilihat pada Tabel 7.

Tabel 7. *Classification Data Testing Random Forest*

Class	*Accuracy*	*Precission*	*Recall*	*F1-score*
0		0.860	0.902	0.881
1	0.8334	0.765	0.684	0.722

Hasil pada Tabel 7 menunjukkan bahwa algoritma *random forest* mencapai akurasi sebesar 0,8334 dengan nilai rata-rata recall dan F1-score yang lebih tinggi dibandingkan algoritma *logistic regression* dan *decision tree*, sementara perbedaan nilai *precision* hanya sebesar 0,0095 yang tetap mengindikasikan konsistensi klasifikasi yang lebih baik, sehingga memperkuat posisi Random Forest sebagai algoritma terbaik untuk klasifikasi dalam penelitian ini.

Berdasarkan hasil pengujian menggunakan data *testing*, algoritma *random forest* memberikan performa evaluasi terbaik di antara ketiga algoritma yang telah diuji. Selain itu, hasil *confusion matrix* untuk algoritma tersebut disajikan secara rinci pada Gambar 3 sebagai gambaran performa klasifikasi yang telah dihasilkan



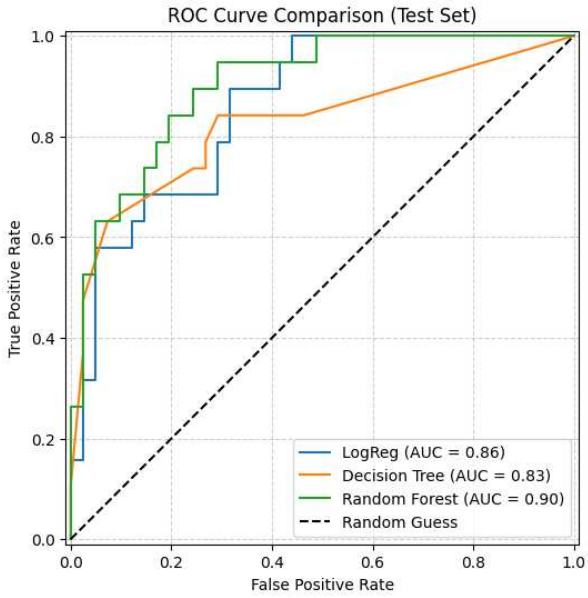
Gambar 3. *Confusion Matrix Random Forest*

Pada Gambar 3 ditampilkan hasil *confusion matrix* algoritma *random forest* dengan pengujian menggunakan data *testing*. Algoritma *random forest* berhasil memprediksi dengan benar sebanyak 38 data untuk kelas 0 (pasien yang hidup) dan 13 data untuk kelas 1 (pasien yang tidak hidup). Namun, masih terdapat 3 data kelas 0 yang diprediksi salah sebagai *class* 1, serta 6 data *class* 1 yang diprediksi salah sebagai *class* 0. Hasil ini menggambarkan tingkat akurasi model dalam mengklasifikasikan kondisi pasien berdasarkan variabel-variabel medis yang telah dianalisis dan menunjukkan kemampuan model secara efektif.

Setelah melakukan evaluasi terhadap *confusion matrix* pada algoritma *random forest*, tahap berikutnya adalah

DOI: https://doi.org/10.38204/tematik

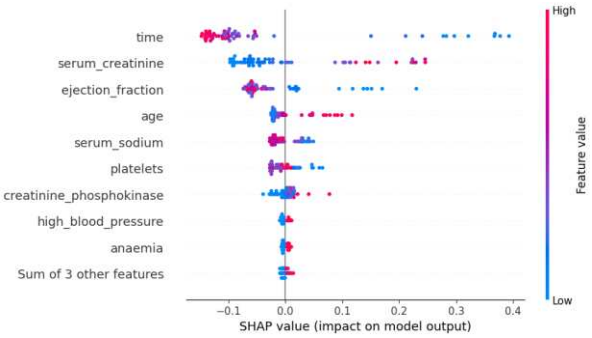
melakukan perbandingan kurva ROC (*Receiver Operating Characteristic*) untuk ketiga algoritma machine learning yang digunakan. Tujuannya untuk memperoleh visualisasi yang komprehensif mengenai kinerja ketiga model dalam membedakan positif *class* dan negatif *class* pada dataset *testing*. Adapun hasil komparasi dapat dilihat pada Gambar 4. Berdasarkan gambar 4 *ROC curve comparison*, dapat dilihat bahwa algoritma *random forest* memiliki nilai *AUC* tertinggi sebesar 0,90, diikuti oleh *logistic regression* dengan *AUC* sebesar 0,86, dan *decision tree* dengan *AUC* sebesar 0,83. Nilai *AUC* mengukur kemampuan model dalam membedakan antara *class* positif dan negatif, sehingga semakin besar nilai *AUC*, semakin baik performa model dalam klasifikasi. *Curve random forest* berada di atas *curve* algoritma lainnya, menunjukkan bahwa *random forest* paling efektif dalam memprediksi target *class* dibandingkan *logistic regression* dan *decision tree* pada data *testing*. Selain itu, garis putus-putus menunjukkan prediksi acak sebagai *baseline*, yang jauh lebih rendah dibandingkan kedua algoritma tersebut.



Gambar 4. *ROC Curve Comparison*

3.4. Menerapkan *Interpretability*

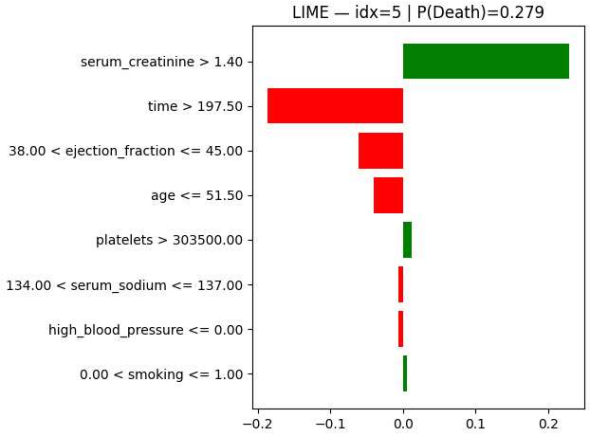
Tahap selanjutnya dalam penelitian ini adalah penerapan *interpretability* untuk menjelaskan hasil evaluasi yang telah diperoleh. Langkah ini dilakukan agar proses pembentukan *interpretability* dapat dilaksanakan dengan baik, sehingga memungkinkan pemahaman yang lebih mendalam mengenai bagaimana algoritma *random forest* membuat prediksi serta variabel apa saja yang memengaruhi keputusan tersebut. Dengan penerapan *interpretability* tidak hanya menghasilkan prediksi yang akurat, tetapi juga transparan. *Interpretability* yang digunakan dalam penelitian ini adalah metode *SHAP* dan metode *LIME*.Metode *SHAP* digunakan karena memiliki tujuan untuk memberikan hasil interpretasi yang jelas dan transparan terhadap prediksi model *machine learning*. Adapun hasil analisis *interpretability* metode *SHAP* pada Gambar 5.



Gambar 5. Analisis Hasil *SHAP*

Berdasarkan hasil analisis *SHAP* pada Gambar 5, setiap titik dengan warna berbeda merepresentasikan satu *instance* data yang menunjukkan besarnya kontribusi masing-masing variabel terhadap prediksi model. Warna merah pada titik menunjukkan nilai variabel yang tinggi, sedangkan warna biru menunjukkan nilai variabel yang rendah. Analisis ini mengindikasikan bahwa variabel *time* memiliki pengaruh paling signifikan terhadap prediksi model, diikuti oleh serum *creatinine*, *ejection fraction*, dan *age*. Hasil dengan nilai positif mengartikan bahwa nilai variabel tersebut meningkatkan risiko gagal jantung, sementara hasil dengan nilai negatif menunjukkan menurunnya risiko gagal jantung. Dengan demikian, hasil menggunakan *SHAP* memberikan pemahaman mendalam mengenai peran variabel di dalam *dataset* serta perubahan nilai variabel dapat meningkatkan atau mengurangi prediksi risiko gagal jantung.

Metode *LIME* digunakan karena memiliki tujuan untuk menjelaskan hasil dari masing – masing variabel di setiap pasien secara lokal. Adapun hasil analisis *interpretability* metode *LIME* pada Gambar 6.



Gambar 6. Analisis Hasil *LIME*

DOI: https://doi.org/10.38204/tematik

Hasil analisis *LIME* pada Gambar 6 dapat dilihat berdasarkan data pasien nomor 5 dari sumber dataset yang telah digunakan. Hasil analisis *LIME* menunjukkan dua warna bar yang merepresentasikan pengaruh setiap variabel terhadap prediksi gagal jantung. Bar merah pada variabel menandakan kontribusi cenderung menurun dalam memprediksi kematian, sedangkan bar hijau menandakan kontribusi cenderung meningkat dalam memprediksi kematian. Variabel *serum creatinine* > 1.00 memiliki hasil yang signifikan dalam meningkatkan hasil memprediksi kematian. Sebaliknya, variabel seperti *time* > 107.50, *ejection fraction* <= 38.00, dan *age* <= 75.00 justru menurunkan kemungkinan memprediksi kematian. Kontribusi variabel *platelets* > 353000.00, *serum sodium* <= 137.00, *high blood pressure* <= 0.00, dan *smoking* <= 0.00 sangat kecil terhadap prediksi gagal jantung dengan metode *LIME* pada kasus ini.

4. Kesimpulan

Penelitian ini berhasil menemukan algoritma *machine learning* terbaik dalam memprediksi gagal jantung. Algoritma *random forest* dipilih sebagai algoritma terbaik berdasarkan hasil evaluasi yang sudah dilakukan menggunakan data training dan data *testing*, hasil *accuracy* algoritma *random forest* menunjukkan nilai sebesar 0.8334 secara keseluruhan diikuti dengan nilai rata-rata *precission* sebesar 0,8125, nilai rata-rata *recall* sebesar 0,793, nilai rata-rata *f1-scorenya* sebesar 0,8015 dan nilai *UAC*nya sebesar 0.90. hasil tersebut menunjukkan performa yang seimbang dan stabil di kedua *class* yaitu *class 0* dan *class* 1. Nilai metrik yang optimal ini menegaskan bahwa *random forest* efektif dalam mengatasi data yang kompleks dan *imbalanced*, serta mampu memberikan keseimbangan antara tingkat deteksi yang benar dan kesalahan prediksi, sehingga menghasilkan model yang lebih andal dan robust dibandingkan algoritma *logistic regression* dan *decision tree*. Setelah model terbaik diperoleh, diterapkan metode interpretabilitas *SHAP* dan *LIME* untuk memahami faktor-faktor yang memengaruhi prediksi. Hasil *SHAP* menunjukkan bahwa variabel *time*, *serum creatinine*, *ejection fraction*, dan *age* memiliki pengaruh terbesar terhadap risiko gagal jantung. Temuan ini signifikan dalam konteks medis karena variabel-variabel tersebut memang mencerminkan progresivitas kondisi jantung, fungsi ginjal, dan kapasitas pompa ventrikel yang secara klinis relevan dalam menentukan prognosis pasien. Sementara itu, hasil *LIME* pada kasus individual (misalnya pasien nomor 5) menegaskan bahwa *serum creatinine > 1.00* menjadi indikator lokal yang meningkatkan prediksi risiko kematian. Pendekatan ini memperlihatkan bagaimana faktor tertentu memengaruhi keputusan model pada level pasien. Dengan pendekatan ini, penelitian ini memberikan kontribusi penting dengan menghadirkan model prediksi gagal jantung yang tidak hanya akurat tetapi juga transparan. Integrasi *SHAP* dan *LIME* memungkinkan tenaga medis memahami alasan di balik prediksi model, sehingga hasil yang diperoleh lebih mudah diinterpretasikan untuk mendukung keputusan klinis. Penelitian ini diharapkan dapat menjadi landasan dalam pengembangan sistem pendukung keputusan berbasis *machine learning* yang lebih dapat dipertanggungjawabkan di bidang kardiologi .

Daftar Rujukan

[1] R. R. Sanni and H. S. Guruprasad, "Analysis of performance metrics of heart failured patients using Python and machine learning algorithms," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 233–237, 2021, doi: https://doi.org/10.1016/j.gltp.2021.08.028.

[2] V. Venkat, H. Abdelhalim, W. DeGroat, S. Zeeshan, and Z. Ahmed, "Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine," *Genomics*, vol. 115, no. 2, p. 110584, 2023, doi: https://doi.org/10.1016/j.ygeno.2023.110584.

[3] A. C. A. Rao, M. Neshat, and P. J. Scott, "Enhancing Clinical Decision-Making in Heart Failure: SHAP-Based Interpretability of Length of Stay Predictions Using a LightGBM Machine Learning Model," *Heart Lung Circ*, vol. 34, pp. S492–S494, 2025, doi: https://doi.org/10.1016/j.hlc.2025.06.618.

[4] W. Wang *et al.*, "Potential diagnostic biomarkers in heart failure: Suppressed immune-associated genes identified by bioinformatic analysis and machine learning," *Eur J Pharmacol*, vol. 986, p. 177153, 2025, doi: https://doi.org/10.1016/j.ejphar.2024.177153.

[5] A. Newaz, N. Ahmed, and F. Shahriyar Haq, "Survival prediction of heart failure patients using machine learning techniques," *Inform Med Unlocked*, vol. 26, p. 100772, 2021, doi: https://doi.org/10.1016/j.imu.2021.100772.

[6] G. de M. C. Gondim *et al.*, "Reliability, internal consistency, and validity of the World Health Organization disability assessment schedule (WHODAS) 2.0 among adults with heart failure," *Heart & Lung*, vol. 70, pp. 30–35, 2025, doi: https://doi.org/10.1016/j.hrtlng.2024.11.003.

[7] K. Nakajima, T. Nakata, T. Doi, H. Tada, and K. Maruyama, "Machine learning-based risk model using 123I-metaiodobenzylguanidine to differentially predict modes of cardiac death in heart failure," *Journal of Nuclear Cardiology*, vol. 29, no. 1, pp. 190–201, 2022, doi: https://doi.org/10.1007/s12350-020-02173-6.

[8] B. J. Van Essen *et al.*, "Obesity and inactivity cluster the strongest risk factor for the development of heart failure in a population-based study," *Int J Cardiol*, p. 133914, 2025, doi: https://doi.org/10.1016/j.ijcard.2025.133914.

[9] C. Mann, E. Braunwald, and T. A. Zelniker, "Diabetic cardiomyopathy revisited: The interplay between diabetes and heart failure," *Int J Cardiol*, vol. 438, p. 133554, 2025, doi: https://doi.org/10.1016/j.ijcard.2025.133554.

[10] V. Goel *et al.*, "Utilising Machine Learning to Predict Heart Failure Readmission Following Emergency Department Presentation," *Heart Lung Circ*, vol. 34, p. S505, 2025, doi: https://doi.org/10.1016/j.hlc.2025.06.637.

[11] R. T. Levinson, C. Paul, A. D. Meid, J.-H. Schultz, and B. Wild, "Identifying Predictors of Heart Failure Readmission in Patients From a Statutory Health Insurance Database: Retrospective Machine Learning Study," *JMIR Cardio*, vol. 8, 2024, doi: https://doi.org/10.2196/54994.

[12] Md. M. Ali *et al.*, "A machine learning approach for risk factors analysis and survival prediction of Heart Failure

DOI: https://doi.org/10.38204/tematik

patients," *Healthcare Analytics*, vol. 3, p. 100182, 2023, doi: https://doi.org/10.1016/j.health.2023.100182.

[13] D. XU *et al.*, "Machine Learning for Proteomic Risk Scores in Heart Failure," *J Card Fail*, vol. 29, no. 11, pp. 1583–1585, 2023, doi: https://doi.org/10.1016/j.cardfail.2023.08.023.

[14] A. Naseri, D. Tax, P. van der Harst, M. Reinders, and I. van der Bilt, "Data-efficient machine learning methods in the ME-TIME study: Rationale and design of a longitudinal study to detect atrial fibrillation and heart failure from wearables," *Cardiovasc Digit Health J*, vol. 4, no. 6, pp. 165–172, 2023, doi: https://doi.org/10.1016/j.cvdhj.2023.09.001.

[15] A. S. Sunge, Amali, A. T. ZY, D. K. Pramudito, A. Badruzzaman, and Purwanto, "The model interpretability on SHAP and comparison classification selection feature for heart disease prediction," *Procedia Comput Sci*, vol. 245, pp. 210–219, 2024, doi: https://doi.org/10.1016/j.procs.2024.10.245.

[16] X. Peng *et al.*, "Development and validation of risk stratification and shared decision-making tool for catheter ablation for atrial fibrillation in patients with heart failure: a multicentre cohort study," *EClinicalMedicine*, vol. 83, p. 103219, 2025, doi: https://doi.org/10.1016/j.eclinm.2025.103219.

[17] K. Wang *et al.*, "Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP," *Comput Biol Med*, vol. 137, p. 104813, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104813.

[18] D. Scrutinio *et al.*, "Prediction of mortality in heart failure by machine learning. Comparison with statistical modeling," *Eur J Intern Med*, vol. 133, pp. 106–112, 2025, doi: https://doi.org/10.1016/j.ejim.2025.01.020.

[19] M. Gaudin, S. Kaur, P. Sharma, and R. Kumar, "A Comparative Analysis of Machine Learning Models for the Classification of Heart Failure Patients in the Intensive Care Unit," *Recent Advances in Electrical and Electronic Engineering*, vol. 18, no. 7, pp. 834–849, 2025, doi: https://doi.org/10.2174/0123520965312805240506113451.

[20] H. Yao, J. R. Golbus, J. Gryak, F. D. Pagani, K. D. Aaronson, and K. Najarian, "Identifying potential candidates for advanced heart failure therapies using an interpretable machine learning algorithm," *The Journal of Heart and Lung Transplantation*, vol. 41, no. 12, pp. 1781–1789, 2022, doi: https://doi.org/10.1016/j.healun.2022.08.028.

[21] J. Lamp *et al.*, "Characterizing advanced heart failure risk and hemodynamic phenotypes using interpretable machine learning," *Am Heart J*, vol. 271, pp. 1–11, 2024, doi: https://doi.org/10.1016/j.ahj.2024.02.001.

[22] A. Cai, Y. Zhu, S. A. Clarkson, and Y. Feng, "The Use of Machine Learning for the Care of Hypertension and Heart Failure," *JACC: Asia*, vol. 1, no. 2, pp. 162–172, 2021, doi: https://doi.org/10.1016/j.jacasi.2021.07.005.

[23] S. Y. Jang *et al.*, "Mortality Prediction in Patients With or Without Heart Failure Using a Machine Learning Model," *JACC: Advances*, vol. 2, no. 7, p. 100554, 2023, doi: https://doi.org/10.1016/j.jacadv.2023.100554.

[24] Alda Amorita Azza, Asep Id Hadiana, and Agus Komarudin, "Klasifikasi Aktivitas Pengguna yang Berpotensi Menyebabkan Kebocoran Informasi Sensitif Menggunakan Algoritma Random Forest," *TEMATIK*, vol. 12, no. 1, pp. 41–49, Jun. 2025, doi: 10.38204/tematik.v12i1.2325.

DOI: https://doi.org/10.38204/tematik