

Optimization of Heart Failure Risk Prediction Using Random Forest Classifier Algorithm

M Nabil Fadhlurrahman

Department of Informatics, Dinamika Bangsa University, Jambi, Indonesia

E-mail: mnabilf81@gmail.com

Eko Arip Winanto

Department of Computer Engineering, Dinamika Bangsa University, Jambi, Indonesia

E-mail: ekoaripwinanto@gmail.com

*Corresponding Author

Received: 22 May, 2025; Accepted: 2 June, 2025; Published: 30 June, 2025

Abstract: This study discusses the optimization of heart failure prediction using the Random Forest Classifier algorithm with a focus on feature selection marked by a threshold and the number of features used. The results of the analysis show that the right threshold has a significant effect on model performance. At a threshold of 0.02, the model achieves the best performance with the highest accuracy, precision, and F1-score values. However, increasing the threshold above 0.08 causes a gradual decrease in model performance. In addition, the number of features used also affects the prediction results, where the right combination of features can increase the effectiveness of the classification. Therefore, this study emphasizes the importance of optimizing thresholds and feature selection in building more accurate and efficient prediction models

Keywords: Optimization, Prediction, Heart Failure, Algorithm, Random Forest Classifier

I. Introduction

Penyakit kardiovaskular, termasuk gagal jantung, merupakan penyebab kematian tertinggi di dunia. Data WHO menunjukkan 17,8 juta kematian global tiap tahun[1], Sementara di Indonesia angkanya mencapai 650.000 kasus per tahun[2]. Gagal jantung, yang sering berkembang dari kondisi akut ke kronis, dipicu oleh gangguan fungsi ventrikel kiri akibat hipertensi, kardiomiopati, atau penyakit jantung koroner. Deteksi dini menjadi kunci untuk mengurangi mortalitas melalui intervensi yang tepat waktu.

Machine learning telah mengubah pendekatan prediksi penyakit melalui analisis data klinis, demografis, dan gaya hidup, sebagaimana dipaparkan [3]. Khususnya pada CVD, teknologi ini mendukung deteksi dini dengan mengidentifikasi keterkaitan kompleks antarfaktor risiko seperti tekanan darah tinggi, kolesterol, dan riwayat keluarga, yang kerap tak terdeteksi oleh cara tradisional. Penelitian membuktikan bahwa ML meningkatkan ketepatan prediksi, mempercepat intervensi, dan meringankan beban sistem kesehatan dengan menentukan pasien berisiko tinggi lebih awal [4]. Hal ini krusial mengingat melonjaknya kasus CVD, dan model prediktif berperan besar dalam menekan angka kematian [5].

Selain meningkatkan akurasi, integrasi *Machine Learning* dalam diagnosis penyakit jantung juga mempercepat proses deteksi dini, yang krusial untuk intervensi medis yang sukses. Penelitian menunjukkan bahwa model *Machine Learning* yang dilatih pada dataset besar dapat memprediksi risiko penyakit jantung hingga lima tahun ke depan dengan performa yang lebih baik dibandingkan model risiko konvensional [6]. Dalam analisis citra medis, seperti *echocardiogram*, *Machine Learning* memungkinkan deteksi otomatis kelainan struktural jantung, seperti *kardiomiopati*, dengan tingkat keakuratan yang mendekati kemampuan ahli kardiologi [7]. Meskipun demikian, validasi model pada populasi yang beragam tetap menjadi tantangan karena keterbatasan akses data dan isu privasi pasien [8].

Penggunaan *Machine Learning*, terutama Algoritma *Random Forest* dengan seleksi fitur, menawarkan kemajuan dalam prediksi dan penanganan CVD [5]. Metode ini membantu tenaga kesehatan mendiagnosis secara lebih akurat dan cepat, sehingga meningkatkan hasil pengobatan pasien serta efisiensi sumber daya medis. Penelitian ke depan perlu mengembangkan model canggih yang mengintegrasikan data genetik, informasi dari perangkat wearable, dan aspek psikologis, dengan tetap menjaga kejelasan dan kemudahan penggunaan dalam praktik klinis[9]. Langkah ini dapat mendukung pengobatan yang lebih personal dan efektif, sekaligus menekan dampak CVD di seluruh dunia.

Beberapa Penelitian terkait dengan *Machine Learning* seperti yang dilakukan oleh [10] dengan mengevaluasi empat algoritma pembelajaran *Machine Learning* *Random Forest*, *Support Vector Machine* (SVM), *K-Nearest*

Neighbors (KNN) dan Regresi Logistik menghasilkan akurasi terbaik menggunakan *Random Forest* dengan akurasi 87.7%. Sedangkan [11] menggunakan algoritma *Random Forest* mengalami peningkatan dari 82,6% menggunakan Teknik Optimasi *K-Fold* dan *GridSearchCV* menjadi 85%. Berdasarkan beberapa penelitian diatas bahwa *Machine Learning* dapat digunakan dalam prediksi gagal jantung dengan tingkat akurasi yang baik. Maka peneliti akan melakukan penelitian prediksi gagal jantung menggunakan metode *Feature Selection*.

Seleksi fitur didefinisikan sebagai teknik pengurangan dimensi yang bertujuan memilih *subset* fitur relevan dari fitur asli dengan menghilangkan fitur yang tidak relevan, berlebihan, atau *noise*. Menurut [12] seleksi fitur bertujuan meningkatkan performa pembelajaran dengan akurasi lebih tinggi, biaya komputasi lebih rendah, dan interpretabilitas model yang lebih baik. Sementara itu, [13] menyoroti peran seleksi fitur dalam penambangan data besar, menangani tantangan data berdimensi tinggi untuk meningkatkan akurasi model. [14] memberikan gambaran menyeluruh, menyatakan bahwa seleksi fitur mengurangi dimensi fitur dan kompleksitas komputasi dengan menghilangkan fitur yang tidak relevan dan berlebihan, meningkatkan performa pada berbagai model seperti klasifikasi, regresi, dan klastering

Feature Selection telah menjadi fokus utama dalam penelitian Machine Learning, seperti yang diulas dalam [15] yang menyoroti perkembangan teknik Feature Selection selama lebih dari dua dekade. Pada buku [16] memberikan panduan praktis tentang penerapan metode seleksi fitur, seperti *filter*, *wrapper*, dan *embedded*, untuk meningkatkan performa model dengan mengurangi kompleksitas dan mencegah *overfitting*. Dalam konteks medis, [17] menekankan pentingnya seleksi fitur dalam prediksi risiko penyakit, khususnya dengan mengekstraksi fitur informatif dan menghilangkan fitur *noise*. Sementara itu, pada penelitian [18] menunjukkan bahwa seleksi fitur dapat secara signifikan meningkatkan akurasi model dalam prediksi penyakit jantung, sekaligus mengurangi waktu komputasi dan risiko *overfitting*.

Machine learning menawarkan solusi melalui kemampuannya mengolah data kompleks. Algoritma *Random Forest* dengan potensi optimasi model dengan teknik *feature selection*-seleksi fitur untuk memilih variabel paling signifikan—belum banyak dieksplorasi. Padahal, pendekatan ini dapat meningkatkan akurasi dan efisiensi model dengan mengurangi redundansi data. Berdasarkan urgensi tersebut, penelitian ini bertujuan mengoptimalkan prediksi gagal jantung menggunakan *Random Forest Classifier* yang diintegrasikan dengan *feature selection*. Hasilnya diharapkan menghasilkan model yang lebih akurat dan andal, sehingga mendukung tenaga kesehatan dalam identifikasi risiko pasien secara dini. Dengan demikian, intervensi pencegahan dan penanganan dapat dilakukan lebih cepat, menekan angka kematian serta dampak sosial-ekonomi akibat penyakit ini

2. Method

2.1. Experiment Setup

Penelitian ini bertujuan untuk mengidentifikasi fitur-fitur apa saja yang mempengaruhi review dan memetakan point-point penting agar dapat menarik kesimpulan dari data dengan efektif. Dalamaaaaaa mencapai tujuan tersebut, peneliti menggunakan algoritma Random Forest, Principal Component Analysis dan Feature Selection sebagai alat menghitung data. Berikut beberapa tahap penelitian yang perlu dilakukan akan diuraikan secara rinci dalam alur penelitian yang ditampilkan dalam Fig 1.

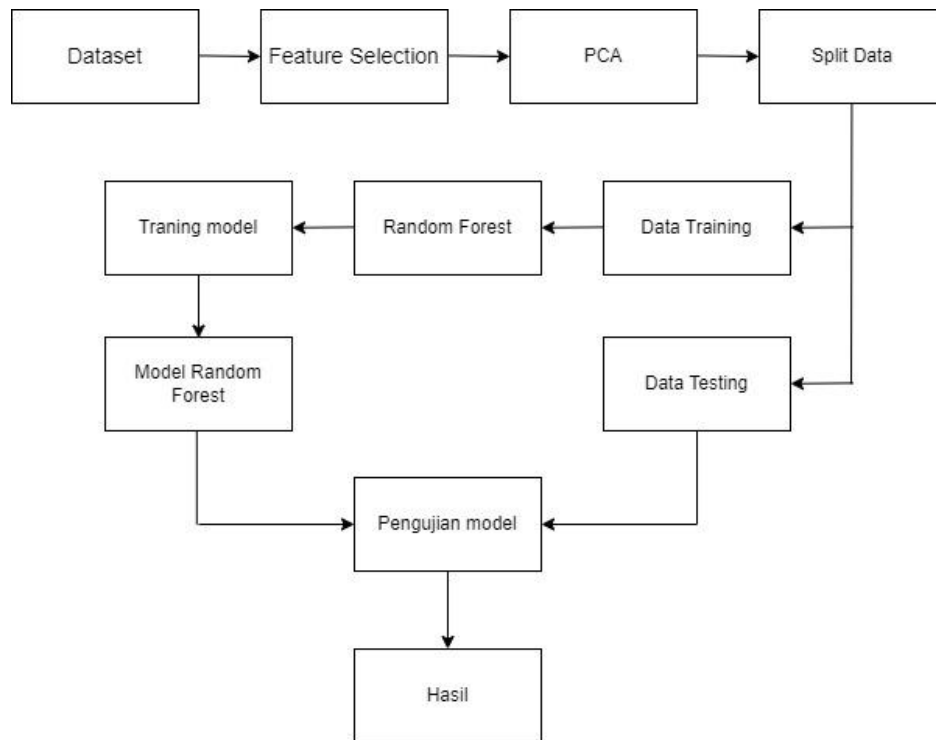


Fig.1. Experiment Setup.

Fig 1 merupakan diagram eksperimen yang menunjukkan tahapan-tahapan yang akan dilakukan dalam penelitian ini dengan melibatkan tiga fase tahapan terpisah, yaitu :

- Mengatur dataset dan mengekstrak fitur-fiturnya dengan Feature selection, lalu akan dibuat iterasi dimana akan membagi jumlah fitur dan data training yang digunakan. Lalu dataset tersebut menjadi dua bagian: data training dan data testing.
- Pada tahap ini, model akan dilatih menggunakan algoritma Random Forest untuk memperoleh hasil yang akan digunakan dalam pengujian model.

Pada tahap akhir, model yang telah dilatih akan diuji, dan tingkat keberhasilan peningkatan dalam akurasi, presisi, serta recall akan dihitung.

2.2 Dataset

Dataset yang digunakan pada penelitian ini didapatkan dari dataset publik penyakit gagal jantung yang ada di *Kaggle.com*, dataset ini merupakan gabungan 5 dataset yang digunakan dari Cleveland, Hungarian, Switzerland, Long Beach VA and Stalog pasien penyakit gagal jantung. Dataset ini memiliki 11 fitur klinis untuk memprediksi gagal jantung yang diantaranya, *Age* (umur), *Sex* (Jenis Kelamin), *ChestPainType* (Jenis Nyeri Dada), *RestingBP* (Tekanan darah saat istirahat), *Cholestrol*, *FastingBs* (Gula darah saat tidak makan), *RestingECG* (Elektrokardiografi saat istirahat), *MaxHR* (Denyut Jantung Maksimal), *ExerciseAngina* (Nyeri saat olahraga), *OldPeak* (Indikator stress), *ST_Slope* (Kemiringan segmen ST) dan *HeartDisease* (Indikator Penyakit Jantung). Dengan memahami dataset berikut dapat dimanfaatkan untuk memperoleh informasi penting dalam memprediksi gagal jantung.

2.3 Proposed Method

Peneliti menggunakan Random Forest Classifier, sebuah algoritma ensemble learning yang menggabungkan beberapa pohon keputusan selama pelatihan. Output dari Random Forest adalah modus dari kelas (untuk klasifikasi) atau rata-rata prediksi (untuk regresi) dari masing-masing pohon. Random Forest Classifier efektif dalam menangani dataset heart disease karena mampu mengelola sejumlah besar fitur dan menangani missing values dengan baik [3]. Algoritma ini juga membantu dalam mengestimasi pentingnya fitur, yang dapat digunakan bersama teknik feature selection untuk meningkatkan akurasi model secara keseluruhan.

Feature selection adalah proses memilih subset fitur paling relevan dari dataset untuk meningkatkan performa model machine learning, mengurangi kompleksitas, dan mencegah overfitting, terutama pada data berdimensi tinggi seperti deteksi intrusi atau klasifikasi citra [19], [20]. Teknik ini terdiri dari metode filter (berbasis statistik seperti korelasi), wrapper (menggunakan algoritma pencarian seperti recursive feature elimination) [20], dan embedded (terintegrasi dalam algoritma seperti LASSO atau Random Forest) [21]. Penelitian menunjukkan bahwa feature selection meningkatkan akurasi model dan efisiensi komputasi dengan menghilangkan fitur redundan atau tidak relevan, dengan

aplikasi luas dalam prediksi risiko penyakit, analisis genetik, dan klasifikasi data [19], [22], [23].

RFC unggul dalam menangani data dengan noise, fitur kategorikal, dan ketidakseimbangan kelas, menjadikannya ideal untuk aplikasi seperti diagnosis medis, deteksi intrusi, dan analisis genetik[21]. Penelitian terbaru pada 2024 menunjukkan bahwa RFC dengan FS berbasis Mutual Information dan Recursive Feature Elimination mencapai akurasi 92% dalam klasifikasi penyakit kardiovaskular, mengungguli SVM dan KNN. Studi lain pada 2024 menyoroti efektivitas RFC dalam analisis data metabolomik, di mana FS dengan RFC meningkatkan sensitivitas model hingga 20% pada dataset berdimensi tinggi[22]. Kelemahan RFC meliputi kebutuhan komputasi yang tinggi pada dataset besar dan pentingnya tuning parameter seperti jumlah pohon atau kedalaman maksimum[23].

Penelitian [20],[24] menunjukkan bahwa Random Forest dengan feature selection sangat cocok untuk aplikasi seperti prediksi risiko penyakit, klasifikasi citra, dan analisis genomik, karena mampu meningkatkan akurasi dan efisiensi sambil mengurangi risiko overfitting. Terdapat tantangan seperti inkonsistensi hasil antar metode dan ketergantungan pada jenis dataset menuntut pemahaman mendalam dalam memilih pendekatan yang tepat, terutama pada proyek industri atau akademik [25]. Namun, pemilihan metode harus disesuaikan dengan karakteristik dataset dan tujuan proyek, dengan pendekatan hati-hati untuk memastikan hasil optimal.

2.4. Environment Setup

Eksperimen dalam penelitian ini dilaksanakan pada sebuah laptop dengan spesifikasi tertentu, yaitu menggunakan Windows 11, dengan Prosesor Intel Core i5-10300H dengan kecepatan sekitar 2.50GHz serta Ram berkapasitas 16 GB, Beserta alat-alat yang digunakan untuk penelitian meliputi Google Collab, Word, Excel, dan Figma.

3. Result and Discussion

Bagian ini menyajikan hasil eksperimen yang telah dilakukan, termasuk laporan hasil dari langkah-langkah reduksi fitur dan pengujian kinerja algoritma Random Forest. Dalam pembahasan ini, akan dievaluasi efektivitas proses reduksi fitur serta dianalisis kinerja algoritma Random Forest.

3.1 Hasil Pemilihan Fitur

Dalam proses seleksi fitur ini, algoritma Random Forest akan digunakan untuk menangani data dengan dimensi tinggi serta mengurangi risiko overfitting. Tujuan dari pendekatan ini adalah untuk mengurangi beban komputasi dan meningkatkan akurasi serta presisi tanpa kehilangan karakteristik data. Tabel 1 menunjukkan hasil yang dilakukan oleh fitur seleksi dalam menghasilkan tiga fitur penting dalam dataset amazon reviews.

Tabel 1. Hasil pemilihan fitur

Fitur	Skor Fitur
ST_Slope	0.246539
ChestPainType	0.115585
ExerciseAngina	0.113234
Oldpeak	0.108248
MaxHR	0.093691
Age	0.076398
Cholesterol	0.073789
RestingBP	0.067411
Sex	0.043600
RestingECG	0.034010
FastingBS	0.027495

3.2 Penggunaan Random Forest Classifier dengan Feature Selection

Dalam pengujian untuk membuktikan peningkatan akurasi, presisi, dan recall peneliti mencoba menggunakan model Random Forest Classifier dengan iterasi berulang menggunakan data training dan testing yang berbeda juga jumlah fitur digunakan dalam percobaan. Dilakukan pengujian terhadap model ini untuk memastikan hasil dengan mencoba untuk mengatur data training, testing dan jumlah fitur yang digunakan. Hasil pengujian ini disajikan dalam tabel 2 sebagai berikut :

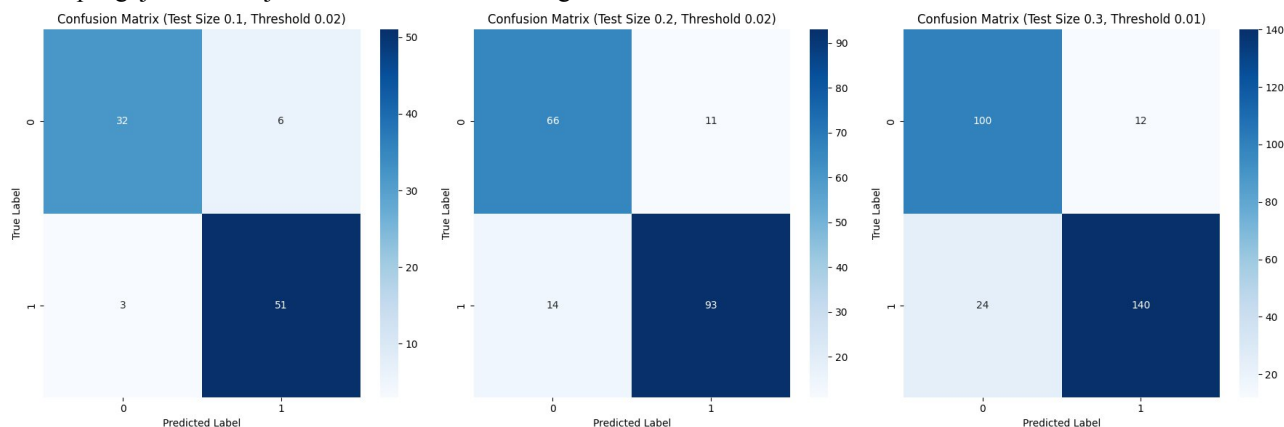
Tabel 2. Hasil Pengujian model Random Forest menggunakan Feature selection

Test Size	Threshold	Num Features	Accuracy	Precision	Recall	F1-Score
0.1	0.01	17	0.880435	0.877193	0.925926	0.900901
0.1	0.02	13	0.902174	0.894737	0.944444	0.918919
0.1	0.03	10	0.847826	0.884615	0.851852	0.867925
0.1	0.04	10	0.847826	0.884615	0.851852	0.867925
0.1	0.05	10	0.847826	0.884615	0.851852	0.867925
0.1	0.06	9	0.836957	0.854545	0.870370	0.862385
0.1	0.07	7	0.836957	0.867925	0.851852	0.859813
0.1	0.08	5	0.793478	0.872340	0.759259	0.811881
0.1	0.09	3	0.739130	0.826087	0.703704	0.760000
0.1	0.10	2	0.760870	0.796296	0.796296	0.796296
0.1	0.11	1	0.793478	0.830189	0.814815	0.822430
0.2	0.01	17	0.836957	0.888889	0.822430	0.854369
0.2	0.02	13	0.864130	0.894231	0.869159	0.881517
0.2	0.03	10	0.826087	0.886598	0.803738	0.843137
0.2	0.04	10	0.826087	0.886598	0.803738	0.843137
0.2	0.05	10	0.826087	0.886598	0.803738	0.843137
0.2	0.06	9	0.798913	0.857143	0.785047	0.819512
0.2	0.07	6	0.766304	0.855556	0.719626	0.781726
0.2	0.08	4	0.771739	0.857143	0.728972	0.787879
0.2	0.09	2	0.782609	0.838384	0.775701	0.805825
0.2	0.10	2	0.782609	0.838384	0.775701	0.805825
0.2	0.11	2	0.782609	0.838384	0.775701	0.805825
0.3	0.01	18	0.869565	0.921053	0.853659	0.886076
0.3	0.02	12	0.855072	0.897436	0.853659	0.875000
0.3	0.03	10	0.836957	0.888889	0.829268	0.858044
0.3	0.04	10	0.836957	0.888889	0.829268	0.858044
0.3	0.05	10	0.836957	0.888889	0.829268	0.858044
0.3	0.06	9	0.811594	0.868421	0.804878	0.835443
0.3	0.07	6	0.815217	0.869281	0.810976	0.839117
0.3	0.08	5	0.804348	0.876712	0.780488	0.825806
0.3	0.09	3	0.782609	0.856164	0.762195	0.806452
0.3	0.10	3	0.782609	0.856164	0.762195	0.806452
0.3	0.11	2	0.764493	0.846154	0.737805	0.788274

Berdasarkan hasil tabel pengujian model Random Forest yang dilakukan dengan data testing dalam rentang 0.1-0.3 disertai dengan jumlah fitur yang telah diencoding dalam rentang threshold 0.01-0.17 menunjukkan bahwa ada peningkatan hasil akurasi, recall, dan presisi yang cukup baik dengan konsistensi hasil yang stabil. Akan tetapi jika jumlah pemilihan fitur diteruskan maka akan menurunkan performa secara signifikan.

3.3 Hasil Confusion Matrix untuk tiap Test Size dengan Threshold Terbaik

Confusion matrix dalam kode ini digunakan untuk mengevaluasi performa model klasifikasi Random Forest Classifier dalam memprediksi penyakit jantung (HeartDisease) menggunakan dataset heart disease, dengan label 0 (Normal) dan 1 (Heart Failure). Matrix ini menunjukkan distribusi prediksi model terhadap data sebenarnya dalam bentuk 2x2, yang terdiri dari True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP). Hasil pengujian ini disajikan dalam Gambar 2 sebagai berikut :

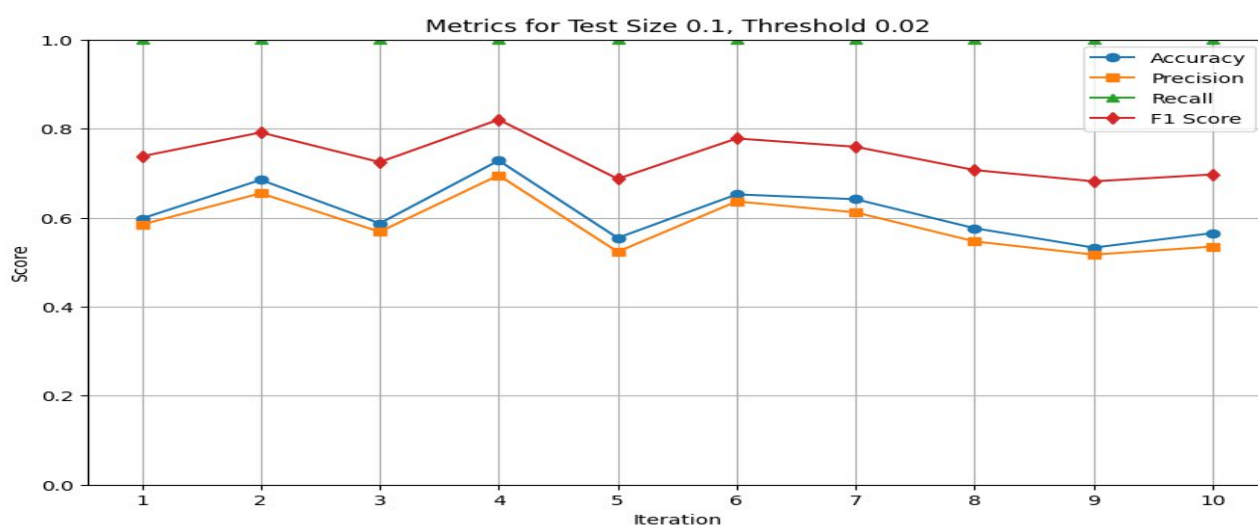


Gambar 2. Confusion Matrix

Dari ketiga Confusion Matrix berikut model dengan Test Size 0.1 dan Threshold 0.02 menunjukkan performa terbaik dibandingkan yang lainnya. Model ini memiliki akurasi tertinggi sebesar 90.7% yang mengindikasikan jumlah prediksi secara keseluruhan lebih unggul. Pada model ini juga memiliki recall tertinggi 82.2% . Secara keseluruhan Confusion Matrix dengan Test Size 0.1 dan Threshold 0.02 memiliki False Negative terendah, Dimana sedikit pasien dengan penyakit jantung tidak terdeteksi dan False Positive untuk kesalahan model untuk memprediksi pasien dengan penyakit jantung.

3.3 Perhitungan Keseimbangan Model

Dilakukan perhitungan untuk 10 iterasi untuk melihat keseimbangan model dalam memprediksi penyakit jantung dengan menggunakan model terbaik dari percobaan. Pada kali ini menghitung 10 percobaan untuk menghitung Akurasi, Presisi, Recall dan F1-score model.



Gambar 3. Akurasi Model setelah di ujicoba 10 kali

Dari gambar tersebut dapat ditemukan bahwa model yang digunakan cukup stabil dalam memperhitungkan pasien penyakit jantung. Model ini memberikan hasil terbaik, dengan akurasi, presisi, recall, dan F1 score yang seimbang serta stabil selama 10 iterasi. Secara keseluruhan, fluktuasi metrik antar iterasi relatif kecil dari yang lainnya.

4. Conclusion

Dari penelitian ini, dapat disimpulkan bahwa penerapan Feature Selection dapat efektif dalam meningkatkan akurasi model Random Forest. Optimasi yang dilakukan telah terbukti pada studi berikut meningkatkan akurasi pada model terbaik dari 88% menjadi 90%. Terdapat beberapa hal yang bisa menjadi pengaruh dalam peningkatan akurasi seperti jumlah data testing dan training juga jumlah fitur yang digunakan, jumlah fitur harusimbang agar tidak mengurangi akurasi yang mana jika semakin sedikit fitur yang dipilih maka performa model akan turun secara signifikan. Penelitian kedepan juga dapat menggunakan cross-validation, hyperparameter untuk meningkatkan model kedepannya.

Acknowledgment

Penelitian didukung oleh Universitas Dinamika Bangsa

References

- [1] World Health Organization (WHO), (2021). "Cardiovascular diseases (CVDs)," World Health Organization (WHO).
- [2] Kementerian Kesehatan Republik Indonesia, "World Heart Day 2023: Use Heart Know Heart," Kementerian Kesehatan Republik Indonesia, 2023. [Online]. Available: <https://www.kemkes.go.id>
- [3] Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. *Diagnostics* (Basel, Switzerland), 14(2), 144. <https://doi.org/10.3390/diagnostics14020144>.
- [4] Dorraki, M, Liao, Z, Abbott, D. et al. (2024), Improving Cardiovascular Disease Prediction With Machine Learning Using Mental Health Data: A Prospective UK Biobank Study. *JACC Adv.* <https://doi.org/10.1016/j.jacadv.2024.101180>
- [5] Bhatt, Chintan & Patel, Parth & Ghetia, Tarang & Mazzeo, Pier Luigi. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms*. 16. 88. 10.3390/a16020088.
- [6] Weng, S. F., et al. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, 12(4), e0174944.
- [7] Madani, A., et al. (2018). Fast and accurate view classification of echocardiograms using deep learning. *npj Digital Medicine*, 1(1), 6.
- [8] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.
- [9] Naser, M.A, et all, (2024). A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. Volume 17(2), Issue 2, <https://doi.org/10.3390/a17020078>
- [10] S. A. Putri, N. Selayanti, M. Kristanaya, M. P. Azzahra, M. G. Navsih, and K. M. Hindrayani, "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung," Seminar Nasional Sains Data, vol. 2024, [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.
- [11] P. Tamba, "PREDIKSI PENYAKIT GAGAL JANTUNG DENGAN MENGGUNAKAN RANDOM FOREST," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [12] Jianyu M, Lingfeng N, "A Survey on Feature Selection", *Procedia Computer Science.*, Volume 91, 2016, Pages 919-926, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.07.111>.
- [13] M. Rong, D. Gong and X. Gao, "Feature Selection and Its Use in Big Data: Challenges, Methods, and Trends," in *IEEE Access*, vol. 7, pp. 19709-19725, 2019, doi: 10.1109/ACCESS.2019.2894366.
- [14] M. R. Islam, A. A. Lima, S. C. Das, M. F. Mridha, A. R. Prodeep and Y. Watanobe, "A Comprehensive Survey on the Process, Methods, Evaluation, and Challenges of Feature Selection," in *IEEE Access*, vol. 10, pp. 99595-99632, 2022, doi: 10.1109/ACCESS.2022.3205618.
- [15] S. Uddin, I. U. Haq, J. Khan, M. A. Salam, dan M. A. Khan, "Feature selection techniques for machine learning: a survey of more than two decades of research," *Knowl. Inf. Syst.*, vol. 65, no. 12, hal. 1–37, Des. 2023, doi: 10.1007/s10115-023-02010-5.
- [16] Data Science Simplified, "Feature selection: Filter method, wrapper method and embedded method," *Data Science Simplified*, Oct. 2019.
- [17] M. Alghamdi, R. Alharbey, dan S. Kocaman, "A review of feature selection methods for machine learning-based disease risk prediction," *Front. Bioinform.*, vol. 2, hal. 927312, Agt. 2022, doi: 10.3389/fbinf.2022.927312.
- [18] N. Singh, P. Singh, dan K. Bhagat, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," *Sci. Rep.*, vol. 13, no. 1, hal. 22805, Des. 2023, doi: 10.1038/s41598-023-49962-w.
- [19] M. R. Karim, T. Rahman, M. S. Islam, and M. A. Rahman, "A review of feature selection methods for machine learning-based disease risk prediction," *Frontiers Bioinformat.*, vol. 2, p. 927312, 2022, doi: 10.3389/fbinf.2022.927312.

- [20] L. Ma, H. Huang, and M. A. Risch, “Integrative feature selection for HDL-C prediction using machine learning,” *BMC Bioinformat.*, vol. 16, no. 1, p. 147, 2015, doi: 10.1186/s12859-015-0565-5.
- [21] M. G. Hafez, M. A. Ghoneim, and H. E. ElDeeb, “Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction,” *Sci. Rep.*, vol. 13, no. 1, p. 49962, (2023), doi: 10.1038/s41598-023-49962-w.
- [22] S. S. Verma, A. Lucas, and T. Zhang, “An ensemble feature selection method for classification,” *Knowl. Inf. Syst.*, vol. 57, no. 2, pp. 307–325, 2018, doi: 10.1007/s10115-018-1184-5.
- [23] N. D. Wahid, A. S. M. Kayes, and M. S. Islam, “Selecting critical features for data classification based on machine learning methods,” *J. Big Data*, vol. 7, no. 1, p. 52, 2020, doi: 10.1186/s40537-020-00327-4.
- [24] R. Iranzad, Y. Liu, Y. Sun, and A. C. Tan, “A review of random forest-based feature selection methods for data science education and applications,” *Int. J. Data Sci. Anal.*, vol. 17, no. 4, pp. 421–439, 2024, doi: 10.1007/s41060-024-00509-w.
- [25] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Syst. Appl.*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/j.eswa.2019.05.028.

Authors’ Profiles



M Nabil Fadhlurrahman lahir di Jambi, Indonesia. Saat ini dia tengah menempuh Program Sarjana Informatika di Universitas Dinamika Bangsa, Indonesia. Memiliki fokus penelitian pada machine learning dan software engineering



Eko Arip Winanto meraih gelar Sarjana di bidang Sistem Komputer dari Universitas Sriwijaya, Indonesia, dan gelar M.Phil di bidang Ilmu Komputer dari Universiti Teknologi Malaysia, Malaysia. Saat ini beliau adalah dosen di Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia. Minat penelitiannya meliputi IoT, pembelajaran mesin, blockchain, dan keamanan jaringan