



Optimasi Fitur Seleksi Random Forest Menggunakan GA Dalam Klasifikasi Data Penyakit Gagal Jantung

Agung Khoeruddin, Fahri Andriansyah Sudrajat, Galuh Purnama,
Iman Kuwangid, Kurnia, Ricky Firmansyah

ARS University

Alamat: Jl. Sekolah Internasional No.1-2 Antapani, Bandung, 40282, Indonesia

Korespondensi penulis: kurniacreativity@email.com

Abstract. Diseases of the heart and blood vessels, such as coronary artery disease (heart attack), cerebrovascular disease (stroke), heart failure (HF), and other pathologies, are collectively referred to as cardiovascular disease (CVD). Globally, around 17 million people a year die from cardiovascular disease, with mortality increasing significantly for the first time in 50 years. Has performed an analysis of the performance of the selection algorithm with case studies predicting the determination of the customer's risk profile. Data mining is the extraction of previously unknown or previously hidden patterns from large databases or data warehouses. This study compares data mining classification models Nave Bayes, Decision Tree, Random Forest, KNN, and SVM to find the most effective model for classifying customer profile data. Later, the most accurate model will be proposed as a replacement model for forecasting the customer's risk profile. As a result, the accuracy value obtained is 82.93% and AUC is 0.896. Then accuracy testing is carried out using the rapidminer application. Testing on rapidminer was carried out with the highest accuracy obtained with an accuracy value of 86.64% and an AUC of 0.880.

Keywords: Random Forest algorithm; datamining; heart failure ; rapidminer.

Abstrak. Penyakit jantung dan pembuluh darah, seperti penyakit arteri koroner (serangan jantung), penyakit serebrovaskular (stroke), gagal jantung (HF), dan patologi lainnya, secara kolektif disebut sebagai penyakit kardiovaskular (CVD). Secara global, sekitar 17 juta orang per tahun meninggal akibat penyakit kardiovaskular, dengan angka kematian meningkat secara signifikan untuk pertama kalinya dalam 50 tahun. Telah melakukan analisis kinerja algoritma selection dengan studi kasus memprediksi penentuan profil risiko nasabah. Data mining adalah ekstraksi pola yang sebelumnya tidak diketahui atau sebelumnya disembunyikan dari database besar atau gudang data. Penelitian ini membandingkan model klasifikasi data mining Nave Bayes, Decision Tree, Random Forest, KNN, dan SVM untuk menemukan model yang paling efektif untuk mengklasifikasikan data profil pelanggan. Nantinya, model yang paling akurat akan diusulkan sebagai model pengganti peramalan profil risiko nasabah. Hasilnya, nilai akurasi yang didapat sebesar 82,93% dan AUC 0,896. Kemudian pengujian akurasi dilakukan dengan menggunakan aplikasi rapidminer. Pengujian pada rapidminer dilakukan dengan akurasi tertinggi didapat dengan nilai akurasi 86,64% dan AUC 0,880.

Kata kunci: algoritma Random Forest; data mining ; gagal jantung ; rapidminer.

LATAR BELAKANG

Penyakit jantung dan pembuluh darah, seperti penyakit arteri koroner (serangan jantung), penyakit serebrovaskular (stroke), gagal jantung (HF), dan patologi lainnya, secara kolektif disebut sebagai penyakit kardiovaskular (CVD). Secara global, sekitar 17 juta orang per tahun meninggal akibat penyakit kardiovaskular, dengan angka kematian meningkat secara signifikan untuk pertama kalinya dalam 50 tahun. Gagal jantung, yang secara khusus terjadi ketika jantung tidak mampu memompa cukup darah ke tubuh, biasanya disebabkan oleh diabetes, tekanan darah tinggi, atau kelainan atau kondisi jantung lainnya[1]. Risiko kematian dapat dikurangi dengan faktor-faktor yang merusak, yang dapat membantu mencegahnya terjadi sejak awal. Memperoleh pengetahuan umum mengenai gejala serangan jantung. Alat yang kurang tepat digunakan hanya untuk mengatur tekanan darah dan aliran darah, serta untuk mempertahankan cara hidup yang sehat. Data dari laboratorium yang belum berfungsi secara efektif dapat digunakan untuk mendeteksi kondisi jantung. Dalam studi ini, detak jantung akan distimulasi dengan penggunaan algoritma klasifikasi untuk penambangan data berdasarkan PSO yang mengukur ketahanan[2]. Gagal jantung merupakan sindrom klinis yang heterogen, yang disebabkan oleh beberapa komorbiditas yang berbeda, dengan penyakit jantung iskemik dan hipertensi sedang kontributor utama untuk perkembangan gagal jantung[3].

KAJIAN TEORITIS

Data mining adalah ekstraksi pola yang sebelumnya tidak diketahui atau sebelumnya disembunyikan dari database besar atau gudang data[4]. Dulu memainkan peran penting dalam berbagai industri, termasuk keuangan, pendidikan, perawatan kesehatan, dan lain - lain[5]. Banyak organisasi menggunakan data mining untuk menganalisis kumpulan data yang sangat besar, untuk meningkatkan proses pengambilan keputusan dan untuk mendapatkan hasil jangka panjang yang lebih baik[6].

Klasifikasi adalah proses melatih atau mempelajari fungsi target untuk memetakan setiap set atribut (fitur) ke salah satu dari sekian banyak label kelas yang tersedia. Ini juga dapat dipahami sebagai proses mengevaluasi objek data untuk menempatkannya di salah satu dari banyak kelas yang tersedia[7] [8].

Pembelajaran mesin, dalam arti luas, mengacu pada proses penerapan model prediktif ke data atau menemukan cluster yang berguna di dalamnya. Tujuan utama pembelajaran mesin adalah mensimulasikan atau meniru kemampuan manusia untuk mengenali pola, meskipun secara objektif, dengan menggunakan komputasi. Pembelajaran mesin sangat membantu ketika kumpulan data yang dianalisis terlalu luas (berisi banyak titik data individual), terlalu rumit (berisi banyak fitur), atau ketika dimaksudkan untuk mengotomatiskan proses analisis data untuk menciptakan proses analisis data yang berulang dan cepat saluran pipa[9].

Beberapa penelitian yang berkaitan dengan analisis kinerja algoritma berbasis fitur seleksi telah dilakukan. Asep Aryfian dan Tjahyono[10] telah melakukan analisis kinerja algoritma selection dengan studi kasus memprediksi penentuan profile resiko nasabah. Penelitian ini membandingkan model klasifikasi data mining Nave Bayes, Decision Tree, Random Forest, KNN, dan SVM untuk menemukan model yang paling efektif untuk mengklasifikasikan data profil pelanggan. Nantinya, model yang paling akurat akan diusulkan sebagai model pengganti peramalan profil risiko nasabah. Hasilnya, nilai akurasi yang didapat sebesar 82.55%.

Beberapa penulis[11] telah melakukan analisis performa algoritma pemilihan dengan studi kasus Analisis Potensi Dana Ritel pada Nasabah PT Bank Tabungan Negara. Untuk mengkategorikan data nasabah menjadi nasabah potensial dan non potensial yang menjadi tujuan dari penelitian ini, perlu dilakukan pemaksimalan potensi dana ritel. Setelah dilakukan optimasi, nilai akurasi naik menjadi 94,47% dari nilai akurasi awal penelitian sebesar 92,17%.

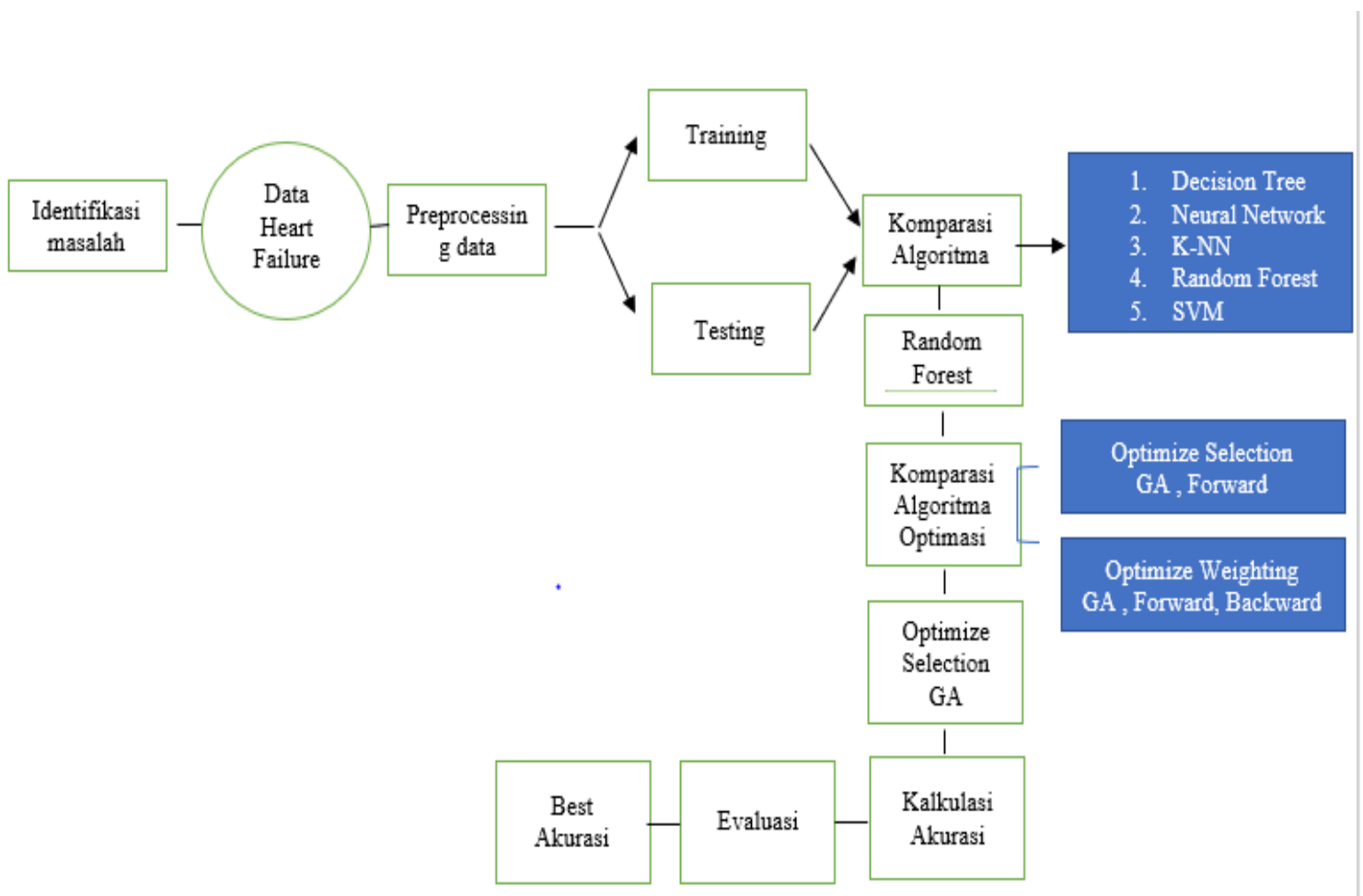
Kami menjelaskan bagaimana meningkatkan algoritma klasifikasi dalam penelitian ini. Menurut penulis, pengoptimalan adalah proses penyelesaian masalah tertentu dalam keadaan yang paling menguntungkan[12]. Penelitian ini menunjukkan bahwa teknik pemilihan fitur optimasi dapat digunakan untuk memanfaatkan hasil klasifikasi untuk dataset yang cukup besar[13].

Penggunaan algoritma genetika (GA) berpotensi memilih sifat yang kurang dominan, maka digunakan untuk mengurangi atribut yang kurang dominan[14]. Untuk mengatasi masalah pengoptimalan di dunia nyata, algoritme genetika mengambil inspirasi dari mode evolusi organisme alami, membuat algoritme evolusi biologis, dan menirunya di komputer. Kemampuan pencarian globalnya memungkinkannya untuk menghindari batasan maksimum lokal[15].

METODE PENELITIAN

2.1 Tahapan Penelitian

Kerangka kerja penelitian ditunjukkan pada Gambar 1. Tahap awal penelitian adalah pengumpulan dataset Heart Failure Clinical melalui UCI Machine Learning. Terdiri dari 299 data dan 13 atribut.



Gambar 1. Kerangka Kerja Penelitian

2.2 Preprocessing Data

Dataset yang telah diambil dari Kaggle masih berupa data mentah tanpa label, oleh karena itu dilakukan tahap pre-processing untuk mendapatkan data bersih agar diproses ke tahap selanjutnya.

2.3 Metode Klasifikasi

Tahap ini menjadi tahap dimana dataset akan dilakukan pengolahan data menggunakan aplikasi RapidMiner dengan membandingkan lima algoritma klasifikasi yang berbeda untuk mencari akurasi dan AUC paling baik diantara kelima algoritma dengan menggunakan operator X-Fold Validation

a. Identifikasi Masalah

Dalam upaya mengklasifikasikan penyakit gagal jantung diperlukan model yang tepat guna membantu para ahli dalam mengklasifikasikan penyakit gagal jantung.

b. Dataset

Penelitian dataset Heart Failure Clinical melalui UCI Machine Learning. Terdiri dari 299 data dan 13 atribut.

c. Komparasi Algoritma

Dalam proses perbandingan algoritma dipilih hasil terbaik dari beberapa algoritma dilakukan pengujian yaitu algoritma Decision Tree, Neural Network, K-NN, SVM, dan Random Forest menggunakan operator X-Fold Validation.

d. Random Forest

Random Forest adalah sebuah metode pembelajaran mesin yang menggabungkan beberapa decision tree untuk membuat sebuah model yang lebih baik. Metode ini digunakan untuk tipe klasifikasi dan regresi. Random Forest digunakan untuk mengatasi masalah overfitting yang sering terjadi pada decision tree tunggal dan juga meningkatkan akurasi model. Random Forest memiliki beberapa kelebihan, antara lain kemampuan untuk meningkatkan performa ketika ada data yang sudah kadaluarsa, tahan terhadap outlier, dan efektif ketika memperluas sekumpulan data. Selain itu, Random Forest memiliki proses untuk memilih fitur terbaik sehingga dapat meningkatkan performa relatif terhadap model klasifikasi. Dengan tambahan fitur ini, Random Forest dapat bekerja dengan data dalam jumlah besar menggunakan parameter yang komprehensif dan efisien[16].

e. X-Fold Validation

X-fold validation merupakan metode statistik yang digunakan guna mempartisi data menjadi bagian training dan testing. Metode ini digunakan untuk membagi data secara berulang menjadi dua bagian, yaitu bagian training dan bagian testing. Setiap data memiliki peluang untuk menjadi data testing[17].

f. Komparasi Algoritma Optimize

Dalam proses pengujian optimize, dilakukan pengujian 2 algoritma optimize yaitu optimize selection dan optimize weighting. Dalam pengujian fitur optimize selection menggunakan 2 algoritma yaitu GA dan Forward, sedangkan dalam pengujian fitur optimize weighting menggunakan 3 algoritma yaitu GA, Backward dan Forward.

g. GA

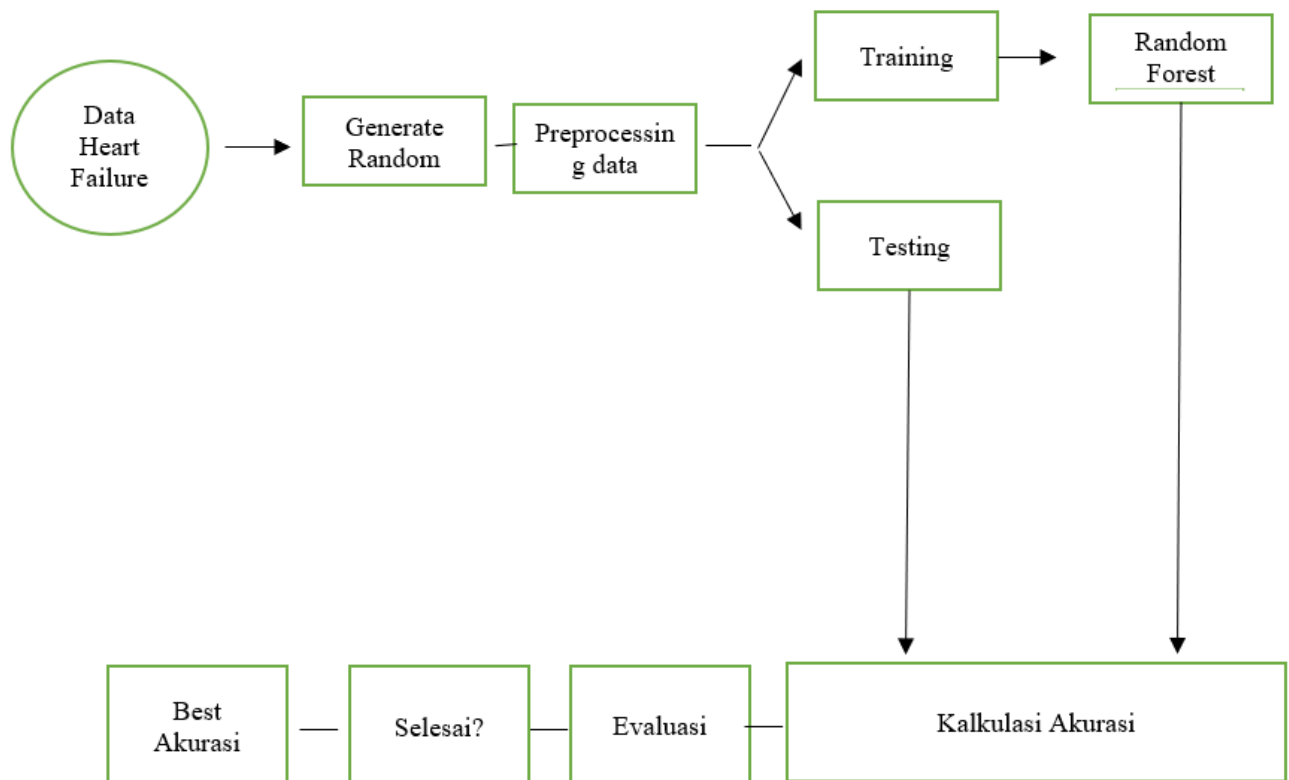
Algoritma evolusioner datang dalam berbagai bentuk, seperti yang diungkapkan oleh sejarah lapangan. Prinsip dasar yang mendasari semua strategi ini adalah sama: mengingat populasi individu di lingkungan dengan sumber daya yang langka, seleksi alam terjadi melalui persaingan untuk mendapatkan sumber daya tersebut[18].

h. Hasil Akhir

Pada tahap ini menampilkan hasil Accuracy dan AUC yang terbaik dan tertinggi yaitu algoritma Random Forest dengan menggunakan fitur optimize selection GA melalui tahap pengolahan pada aplikasi software Rapidminer.

2.4 Metode yang Diusulkan

Dalam penelitian ini, algoritma Random Forest digunakan untuk mengklasifikasikan gagal jantung, dan pendekatan klasifikasi GA disarankan sebagai metode pemilihan fitur. Teknik yang disarankan ditunjukkan pada Gambar 2. Dataset Gagal Jantung dikumpulkan sebagai langkah pertama penelitian. Dataset kemudian diperiksa, dan data kemudian dinormalisasi. Menggunakan pendekatan transformasi-z, dataset dinormalisasi dengan tujuan memblokir data dalam rentang langsung. Data kemudian harus dibagi menjadi set pelatihan dan tes. Pendekatan Random Forest menggunakan data pelatihan untuk membuat model, dan pengujian dataset digunakan untuk membuat nilai yang akurat. Algoritma kemudian harus dibandingkan. Untuk mendapatkan model algoritma yang optimal, algoritma dibandingkan satu sama lain saat mengklasifikasikan. Pemilihan fitur penelitian memanfaatkan GA. Untuk meningkatkan kinerja nilai akurasi klasifikasi, algoritma genetika menghasilkan populasi yang terdiri dari banyak individu yang dipilih dengan cermat dengan nilai yang paling relevan dengan klasifikasi. Selain itu, metode Hutan Acak digunakan untuk mengkategorikan sifat-sifat yang telah dipilih oleh algoritma genetika. Garis besar skema klasifikasi metode yang diusulkan untuk gagal jantung pada Gambar 2. Temuan evaluasi klasifikasi model gagal jantung yang disarankan memiliki nilai paling tinggi karena fitur dioptimalkan menggunakan GA sedemikian rupa sehingga mereka memiliki potensi terbesar untuk mempengaruhi klasifikasi gagal jantung yang dihasilkan oleh algoritma Random Forest.



Gambar 2. Metode yang diusulkan

HASIL DAN PEMBAHASAN

3.1 Pra pengolahan (Pre processing) data

Tahap pertama yang dilakukan pada dataset adalah menerapkan data pre-processing dengan memberikan label pada dataset yang akan digunakan. Pada penelitian ini, proses data pre-processing dilakukan dengan menggunakan aplikasi RapidMiner 9.0.1. Proses ini dilakukan dengan memanfaatkan tools yang ada dalam aplikasi untuk melakukan seluruh tahapan pre-processing. Setelah melakukan tahapan text pre-processing dengan memberikan label, akan dilakukan proses pengujian dengan lima algoritma klasifikasi yang berbeda.

Peneliti melakukan preprocessing data setelah mengumpulkannya. Untuk menentukan apakah ada data yang salah pada saat ini, nilai yang hilang diperiksa pada data. Selanjutnya, hilangkan duplikat untuk memastikan tidak ada dua bagian data yang identik, dan normalkan menggunakan teknik Z-Transformation. Kisaran nilai untuk data

TABEL 1
HASIL NORMALISASI DATASET

Row No.	DEATH_EV...	age	anaemia	creatinine_b...	diabetes	ejection_fra...	high_blood_...	platelets	serum_crea...	serum_sod...	sex	smoking	time
1	1	75	0	582	0	20	1	265000	1.900	130	1	0	4
2	1	55	0	7861	0	38	0	283358.030	1.100	136	1	0	8
3	1	65	0	148	0	20	0	182000	1.300	129	1	1	7
4	1	50	1	111	0	20	0	216000	1.800	137	1	0	7
5	1	65	1	160	1	20	0	327000	2.700	116	0	0	8
6	1	90	1	47	0	40	1	204000	2.100	132	1	1	8
7	1	75	1	248	0	15	0	127000	1.200	137	1	0	10
8	1	50	1	315	1	80	0	454000	1.100	131	1	1	10
9	1	85	0	157	0	65	0	283358.030	1.500	138	0	0	10
10	1	80	1	123	0	35	1	388000	3.400	133	1	1	10
295	0	82	0	61	1	38	1	155000	1.100	143	1	1	270
296	0	55	0	1820	0	38	0	278000	1.200	139	0	0	271
297	0	45	0	2060	1	65	0	742000	0.800	138	0	0	278
298	0	45	0	2413	0	38	0	140000	1.400	140	1	1	280
299	0	50	0	196	0	45	0	395000	1.600	136	1	1	285

3.2 Pengujian dengan 5 Algoritma Klasifikasi

Setelah normalisasi data selesai, selanjutnya membandingkan metode. 5 algoritma yang digunakan dalam penelitian ini dibandingkan menggunakan algoritma Decision Tree, K-Nearest Neighbor (K-NN), Neural Network, SVM, dan Random Forest adalah algoritma yang digunakan. Proses validasi data dilakukan dengan menggunakan X-Fold Validasi yang menghasilkan nilai akurasi dan AUC untuk mengetahui performa dari 5 algoritma. Nilai akurasi dan AUC yang dihasilkan oleh masing-masing algoritma tercantum di bawah ini beserta nilainya masing-masing dari Tabel 2.

TABEL 2
HASIL PERBANDINGAN 5 ALGORITMA

Algoritma	Accuracy	AUC	Validation
Decision Tree	79.92%	0.707	X-Fold Validation
K-NN	62.20%	0.436	X-Fold Validation
SVM	81.60%	0.874	X-Fold Validation
Neural Net	80.24%	0.846	X-Fold Validation
Random Forest	82.93%	0.896	X-Fold Validation

Pada tabel tersebut terdapat hasil perbandingan dari kelima algoritma, pada tabel diatas algoritma Random Forest menjadi algoritma terpilih karena menghasilkan nilai yang tertinggi

dengan akurasi 82,93% dan AUC 0,896 dan adapun hasil nilai terendah dengan akurasi 62,20% dan AUC 0,436 yaitu algoritma K-NN.

TABEL 3
SPLIT RATIO RANDOM FOREST 0,5 – 0,9

Algoritma	Validation	Ratio	Accuracy	AUC
Random Forest	Split	0,5	85,91%	0,842
Random Forest	Split	0,6	86,55%	0,900
Random Forest	Split	0,7	88,89%	0,925
Random Forest	Split	0,8	90,00%	0,918
Random Forest	Split	0,9	93,33%	0,915

Pada tabel diatas terdapat hasil perbandingan algoritma Random Forest menggunakan Split Validation dari ratio 0,5 sampai 0,9

TABEL 4
CONFUSION MATRIX RANDOM FOREST SPLIT 0,5

	TRUE 1	TRUE 0	Class Precision
Pred 1	36	16	69,23%
Pred 0	12	85	87,63%
Class Recall	75,00%	84,16%	

Pada tabel diatas menampilkan hasil Confusion Matrix algoritma Random Forest menggunakan split validation dengan ratio 0,5

3.3 Pengujian dengan kedua fitur optimize

TABEL 5
HASIL PERBANDINGAN 2 ALGORITMA PADA FITUR SELECTION

Algoritma	Accuracy	AUC	Validation
Optimize fitur selection RF			
GA	86.64%	0.880	X-Fold Validation
Greedy Forward	84.62%	0.897	X-Fold Validation
Algoritma Optimize weighting selection RF			
PSO	85.28%	0.902	X-Fold Validation
GA	85.64%	0.882	X-Fold Validation

Pada tabel tersebut terdapat hasil perbandingan kedua fitur optimze, dari masing masing fitur melakukan perbandingan 2 algoritma. Pada fitur optimize selection mendapatkan hasil nilai yang tertinggi dengan akurasi 86,64% dan AUC 0,880 yaitu algoritma GA. Pada fitur optimize weighting selection mendapatkan hasil nilai yang tertinggi dengan akurasi 85,64% dan AUC 0,882 yaitu algoritma GA. Pada pengujian dari kedua fitur optimize yang memiliki nilai tertinggi yaitu fitur optimize selection dengan menggunakan algoritma GA dan adapun fitur atribut yang terpilih yaitu diabetes, ejection frection, high blood pressure, serum sodium, sex, smoking, dan time.

TABEL 6
HASIL ACCURACY, PRECISION, RECALL, AUC RANDOM FOREST + GA SPLIT

Algoritma	Metode	Ratio	Accuracy	Precision	Recall	AUC
Random Forest	GA	0,5	89,93%	91,35%	94,06%	0,939
Random Forest	GA	0,6	89,08%	90,48%	93,83%	0,939
Random Forest	GA	0,7	91,11%	89,55%	98,36%	0,969
Random Forest	GA	0,8	96,67%	97,56%	97,56%	0,974
Random Forest	GA	0,9	100,00%	100,00%	100,00%	1000

Berikut adalah tabel nilai keseluruhan akurasi, presisi, recall, dan AUC dari masing-masing parameter rasio split 0,5 hingga 0,9 algoritma Random Forest berdasarkan Optimize Selection GA.

TABEL 7
HASIL ACCURACY, PRECISION, RECALL, AUC RANDOM FOREST + GA CROSS

Algoritma	Metode	Accuracy	Precision	Recall	AUC
Random Forest	GA	86,64	87,77%	94,14%	0,880

Berikut adalah hasil nilai tertinggi dari perbandingan dua fitur optimize, fitur optimize selection menggunakan algoritma GA yang terpilih menjadi yang terbaik

TABEL 8
PERBANDINGAN HASIL ACCURACY DAN AUC SEBELUM DAN SESUDAH DI OPTIMIZE

Algoritma	Metode	Accuracy	AUC	Validation
Random Forest	-	82,93%	0,896	X-Fold Validation
Random Forest	Fitur Optimize Selection + GA	86,64%	0,880	X-Fold Validation

Berdasarkan tabel diatas terbukti accuracy yang didapat sebelum di optimize yaitu 82,93% dan AUC 0,896 sehingga setelah di optimize accuracu meningkat menjadi 86,64% dan AUC 0,880

KESIMPULAN DAN SARAN

Penyakit kardiovaskular (CVDs) adalah gangguan pada jantung dan pembuluh darah termasuk, penyakit jantung koroner (heart serangan), penyakit serebrovaskular (stroke), gagal jantung (HF), dan jenis patologi lainnya. Berdasarkan tahapan penelitian yang telah dilakukan pada penerapan perbandingan 5 algoritma yaitu Decision Tree, K-NN, NN, SVM dan Random Forest. Random Forest menjadi algoritma terpilih karena menghasilkan nilai tertinggi yaitu 82.93% dan AUC 0,896. Kesimpulan yang diperoleh dari penelitian ini, yaitu, penerapan algoritma Random Forest terhadap data penyakit gagal jantung dilakukan dengan fitur optimize selection menggunakan algoritma GA. Kemudian pengujian akurasi dilakukan dengan menggunakan aplikasi rapidminer. Pengujian pada rapidminer dilakukan dengan akurasi tertinggi didapat dengan nilai akurasi 86,64% dan AUC 0,880 dan fitur atribut yang terpilih yaitu diabetes, ejection fraction, high blood pressure, serum sodium, sex, smoking, dan time.

UCAPAN TERIMA KASIH

Ucapan terimakasih kepada pembimbing dan teman-teman yang sudah banyak membantu dalam pembuatan penelitian ini.

DAFTAR PUSTAKA

- D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med Inform Decis Mak*, vol. 20, no. 1, pp. 1–16, 2020.
- N. A. Widiastuti, S. Santosa, and C. Supriyanto, "Algoritma Klasifikasi data mining naïve bayes berbasis Particle Swarm Optimization untuk deteksi penyakit jantung," *Pseudocode*, vol. 1, no. 1, pp. 11–14, 2014.
- G. D. Lopaschuk, Q. G. Karwi, R. Tian, A. R. Wende, and E. D. Abel, "Cardiac Energy Metabolism in Heart Failure," *Circ Res*, vol. 128, no. 10, pp. 1487–1513, May 2021, doi: 10.1161/CIRCRESAHA.121.318241.

- Z. S. Ageed *et al.*, “Comprehensive survey of big data mining approaches in cloud systems,” *Qubahan Academic Journal*, vol. 1, no. 2, pp. 29–38, 2021.
- W. Haoxiang and S. Smys, “Big data analysis and perturbation using data mining algorithm,” *Journal of Soft Computing Paradigm (JSCP)*, vol. 3, no. 01, pp. 19–28, 2021.
- R. Rastogi and M. Bansal, “Diabetes prediction model using data mining techniques,” *Measurement: Sensors*, vol. 25, p. 100605, 2023.
- R. E. Putri, S. Suparti, and R. Rahmawati, “Perbandingan Metode Klasifikasi Naive Bayes dan k-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012,” *jurnal gaussian*, vol. 3, no. 4, pp. 831–838, 2014.
- J. Jaya Purnama and S. Rahayu, “KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING,” 2022. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, “A guide to machine learning for biologists.”
- A. Arifyan, “Analisis Perbandingan Optimasi berbasis Evolutionary pada Algoritma Klasifikasi Penentuan Profile Resiko Nasabah,” *Techno. Com*, vol. 21, no. 3, pp. 565–578, 2022.
- A. Fauzi and T. Tukiyat, “ANALISIS POTENSI DANA RETAIL PADA NASABAH PT. BANK TABUNGAN NEGARA (PERSERO) TBK DENGAN METODE DECISION TREE DAN NAIVE BAYES BERBASIS OPTIMIZE SELECTION (EVOLUTIONARY (STUDY KASUS: PT. BANK TABUNGAN NEGARA KANTOR KAS SEASON CITY),” *Jurnal Administrasi dan Manajemen*, vol. 9, no. 1, pp. 30–36, 2019.
- S. Muryani and D. Safika, “Rancang Bangun Aplikasi Pemesanan Pada Cantika Catering Berbasis Web,” *Jurnal Teknik Komputer*, vol. 5, no. 2, 2019.
- T. Arifin and A. Herliana, “Optimasi metode klasifikasi dengan menggunakan particle swarm optimization untuk identifikasi penyakit diabetes retinopathy,” *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 2, pp. 77–81, 2018.
- S. Busono, “Optimasi Naive Bayes Menggunakan Algoritma Genetika Sebagai Seleksi Fitur Untuk Memprediksi Performa Siswa,” *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 14, no. 1, pp. 31–40, 2020.
- S. Han and L. Xiao, “An improved adaptive genetic algorithm,” *SHS Web of Conferences*, vol. 140, p. 01044, 2022, doi: 10.1051/shsconf/202214001044.
- S. Devella, Y. Yohannes, and F. N. Rahmawati, “Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 310–320, 2020.
- N. Qomariah, “Sentiment Analysis on Coffee Consumer Perceptions on Social Media Twitter Using Multinomial Naïve Bayes,” *Journal of Intelligent Computing and Health Informatics (JICHI)*, vol. 2, no. 1, pp. 6–11, 2021.
- A. E. Eiben and J. E. Smith, “What is an evolutionary algorithm?,” in *Introduction to evolutionary computing*, Springer, 2015, pp. 25–48.