·Jurnal Komputer dan Informatika
Vol 20 No 2, Oktober 2025: hlm 151 - 159

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN ALGORITMA RANDOM FOREST, SUPPORT VECTOR MACHINE, DAN DECISION TREE

Upie Fitri Nurqalby
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
E-mail: upie.535200068@stu.untar.ac.id

ABSTRAK

Penyakit Kardiovaskular adalah penyakit yang disebabkan oleh penyempitan atau penyumbatan pembuluh darah. Penyakit ini disebabkan oleh tidak berfungsinya jantung dan pembuluh darah sehingga dapat menyebabkan kematian. Menurut WHO, kasus kematian yang disebabkan oleh Penyakit Kardiovaskular sebanyak 18 juta. Penelitian ini bertujuan melakukan klasifikasi untuk memprediksi Penyakit Kardiovaskular dikarenakan penyakit ini cenderung susah yntuk di sembuhkan. Model Klasifikasi yang digunakan untuk memprediksi Penyakit Kardiosvaskular yaitu dengan Algoritma Random Forest, Algoritma Support Vector Machine, dan Algoritma Decision Tree. Dataset yang digunakan adalah 70.000 dengan 13 variabel yaitu id, usia, tinggi, berat badan, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, kolesterol, glukosa, merokok, asupan alkohol, aktivitas fisik, dan kardio (ada tidaknya Penyakit Kardiovaskular). Algoritma Support Vector Machine merupakan algoritma yang memiliki nilai evaluasi tertinggi diantara ketiga algoritma yang digunakan yaitu dengan rata-rata Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72%. Sedangkan Algoritma Decision Tree merupakan algoritma yang memiliki nilai evaluasi terendah diantara ketiga algoritma yang digunakan yaitu dengan rata-rata Accuracy sebesar 70%, Precision sebesar 70%, Recall sebesar 70%, dan F1-Score sebesar 69%.

Kata kunci—Kardiovaskular, Random Forest, Support Vector Machine, Decision Tree, Klasifikasi.

ABSTRACT

Cardiovascular disease is a disease caused by narrowing or blockage of blood vessels. This disease is caused by a malfunction of the heart and blood vessels, which can cause death. According to WHO, there are 18 million deaths caused by cardiovascular disease. This research aims to carry out a classification to predict Cardiovascular Disease because this disease tends to be difficult to cure. The classification model used to predict cardiovascular disease is the Random Forest Algorithm, Support Vector Machine Algorithm, and Decision Tree Algorithm. The dataset used is 70.000 with 13 variables, namely id, age, height, weight, gender, systolic blood pressure, diastolic blood pressure, cholesterol, glucose, smoking, alcohol intake, physical activity, and cardio (presence or absence of Cardiovascular Disease). The Support Vector Machine algorithm is an algorithm that has the highest evaluation value among the three algorithms used, namely with an average Accuracy of 72%, Precision of 72%, Recall of 72%, and F1-Score of 72%. Meanwhile, the Decision Tree Algorithm is an algorithm that has the lowest evaluation value among the three algorithms used, namely with an average Accuracy of 70%, Precision of 70%, Recall of 70%, and F1-Score of 69%.

Keywords— *Cardiovascular, Random Forest, Support Vector Machine, Decision Tree, Classification.*

1 PENDAHULUAN

1.1 Latar Belakang

Penyakit kardiovaskular merupakan kondisi medis yang tidak menular dan dapat menyebabkan kematian sebagai penyebab utama di seluruh dunia. Lebih dari tiga per empat atau sekitar 80% dari kematian akibat penyakit kardiovaskular terjadi di negara dengan tingkat pendapatan rendah hingga menengah, termasuk Indonesia. Penyakit kardiovaskular mencakup gangguan pada jantung dan pembuluh darah, seperti penyakit jantung koroner, gagal jantung, hipertensi, dan stroke. Selain berpotensi fatal, penyakit ini juga dapat mengakibatkan penurunan kualitas hidup terkait kesehatan.

Berdasarkan survei *World Health Organization* (WHO), penyakit kardiovaskular dapat menghentikan fungsi organ tubuh lainnya. Kasus kematian akibat penyakit kardiovaskular mencapai 18 juta orang setiap tahunnya secara global. Meningkatnya angka kasus dan tingginya angka kematian akibat penyakit ini menjadi beban berat bagi sistem perawatan kesehatan di seluruh dunia. Identifikasi penyakit kardiovaskular menjadi sulit karena banyaknya faktor yang berkontribusi, sehingga tantangannya adalah bagaimana melakukan prediksi secara tepat dengan tingkat akurasi yang tinggi.

Dengan perkembangan ilmu pengetahuan dan teknologi informasi, cabang baru dalam bidang komputer yang dikenal sebagai data mining menjadi relevan. Klasifikasi *Machine Learning,* yang melibatkan penempatan objek-objek ke dalam kategori-kategori yang telah ditetapkan sebelumnya, digunakan dalam penelitian ini. Tiga algoritma yang digunakan untuk perbandingan adalah *Random Forest, Support Vector Machine,* dan *Decision Tree*.

1.2 Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan melakukan klasifikasi untuk memprediksi Penyakit Kardiovaskular dikarenakan penyakit ini cenderung susah untuk di sembuhkan. Manfaat dibuatnya topik ini yaitu untuk dapat mengetahui lebih awal mengenai ada atau tidaknya Penyakit Kardiovaskular sehingga dapat ditangani lebih cepat.

2. METODE PENELITIAN

2.1 Studi Literatur

Sri Sumarlinda dan Wiji Lestari menggunakan model prediksi penyakit kardiovaskular dengan algoritma KNN digunakan untuk mengidentifikasi dan memprediksi penyakit kardiovaskular. Algoritma KNN menggunakan jarak Euclidian untuk proses prediksi data latih. Dataset yang digunakan adalah 400 dengan 7 atribut yaitu umur, jenis kelamin, tekanan darah sistolik, kolesterol, talach, oldpeak dan slope. Hasil implementasi algoritma KNN menghasilkan performansi dengan akurasi sebesar 75,75%. Nilai presisinya adalah 76,78%, sedangkan recall menghasilkan 77,14% [1].

Anita Desiani, Muhammad Akbar, Irmeilyana, dan Ali Amran menggunakan metode *support vector machine* dan *naïve bayes* dengan metode latih *percentage split* dan *k-fold cross validation*. Hasil akurasi pengolahan menggunakan Algoritma *Naïve Bayes* adalah sebesar 70% untuk metode latih percentage split dan 71% untuk metode latih *k-fold cross validation*. Kemudian dengan menggunakan algoritma *support vector machine* didapat akurasi 61% untuk metode latih percentage split dan 65% untuk metode latih k-fold validation. Hasil tersebut menunjukkan bahwa algoritma *naïve bayes* dengan metode latih *k-fold validation* cukup baik dalam melakukan klasifikasi penyakit kardiovaskular [2].

Mochammad Anshori menggunakan random forest diimplementasikan untuk membuat model, data dilakukan praproses dengan normalisasi dan menerapkan cross-validation dengan k-fold = 10. Hasil prediksi dengan random forest pada penelitian ini memberikan akurasi sebesar 98%. Akurasi ini lebih tinggi jika dibandingkan dengan penelitian sebelumnya dengan dataset yang sama yaitu 96.75% menggunakan *ensemble method* dan 91.61% dengan *logistic regression*. Dengan dasar tersebut membuktikan bahwa random forest mampu digunakan untuk melakukan prediksi penyakit kardiovaskular [3].

Menurut Wahyu Nugraha, penelitian yang dilakukan menggunakan dataset Kaggle dataset repository (*Heart Failure Prediction*) dapat disimpulkan bahwa model klasifikasi menggunakan XGBOOST rata-rata memperoleh nilai tertinggi baik menggunakan pengukuran accuracy, F1 score maupun AUC. Model klasifikasi SVM memperoleh nilai rata-rata hampir mirip dengan model *Gradient Boosting*. Sedangkan untuk model pengukuran dengan nilai terendah diperoleh dengan model klasifikasi LightGBM [4].

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Menurut Jefri Junifer Pangaribuan, Cathlin Tedja, dan Sentosa Wibowo, Hasil yang diperoleh dari percobaan dan analisis diagnosis penyakit jantung koroner dengan menggunakan metode *Extreme Learning Machine* (ELM) dan Algoritma C4.5 adalah Algoritma C4.5 mempunyai tingkat akurasi yang lebih tinggi dibandingkan dengan ELM dikarenakan Algoritma C4.5 dijabarkan berdasarkan nilai gain tertinggi yang dijadikan nodes dan hasil akhirnya dalam bentuk leaves yang berisikan indikator penentu seseorang penyakit jantung koroner atau tidak [5].

2.2 Metode Pengumpulan Data

Data set penyakit jantung koroner yang digunakan untuk melakukan penelitian ini didapatkan dari Kaggle (https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset). Database yang digunakan dalam penelitian ini berjumlahkan 70.000 dengan menggunakan 13 variabel yaitu id, usia, tinggi, berat badan, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, kolesterol, glukosa, merokok, asupan alkohol, aktivitas fisik, dan kardio (ada tidaknya Penyakit Kardiovaskular).

Dataset berbentuk excel/csv ini memiliki variabel-variabel independent (variabel tidak terikat, biasanya dinotasikan "xx") serta 1 (satu) variabel dependent (variabel terikat, biasanya dinotasikan "yy") sebagai berikut.

Tabel 1 Tabel keterangan Variabel Dependent "y"

Variabel Dependent "y"	
Cardio	Hasil (pasien mengidap penyakit kardiovaskular atau tidak) 0: Tidak 1: Ya

Tabel 2 Tabel keterangan Variabel Independent "x"

Variabel Independent "x"	
Id	Id pasien
Age	Usia pasien
Height	Tinggi pasien
Weight	Berat badan pasien
Gender	Jenis kelamin pasien 1: Wanita 2: Pria
Ap_hi	Tekanan darah sistolik
Ap_lo	Tekanan darah diastolik
Cholestrol	Tingkatan jumlah kolestrol dalam darah 1: Normal 2: Diatas normal 3: Jauh diatas normal
Gluc	Tingkatan jumlah gula dalam darah 1: Normal 2: Diatas normal 3: Jauh diatas normal
Smoke	Pasien merokok atau tidak 0: Tidak 1: Ya
Alco	Pasien mengonsumsi alcohol atau tidak 0: Tidak 1: Ya
Active	Pasien suka melakukan aktivitas fisik atau tidak 0: Tidak 1: Ya

2.2 Metode Klasifikasi

2.2.1 Algoritma Random Forest

Random Forest merupakan suatu algoritma yang sangat fleksibel dan mudah digunakan, memberikan hasil yang baik sebagian besar waktu. Algoritma ini terbentuk dari sejumlah pohon, dimana setiap pohon tumbuh dengan menggunakan berbagai bentuk pengacakan. Operasinya melibatkan pembangunan banyak pohon keputusan selama pelatihan dan menghasilkan kelas yang merupakan modus dari kelas atau prediksi rata-rata dari pohon-pohon individu. Dalam mengevaluasi hasil klasifikasi menggunakan *Random Forest*, beberapa metode evaluasi digunakan seperti Akurasi, *Confusion Matrix, Kappa Statistic, Precision, Recall*, dan tabel laporan klasifikasi. Berikut adalah penjelasan untuk masing-masing metode evaluasi.

a. Akurasi
   Akurasi menggambarkan sejauh mana data benar, dapat diandalkan, dan bersertifikat. Ini mencerminkan tingkat ketepatan, ketelitian, dan kecermatan suatu model. Akurasi memastikan bahwa informasi yang diberikan benar-benar mencerminkan maksud data yang sebenarnya, bebas dari kesalahan dan bias.
b. Confusion matrix
   Confusion Matrix adalah alat pengukuran prediktif yang menggunakan tabel klasifikasi. Tabel ini memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan instance pada kelas yang benar atau salah.

Tabel 3 Klasifikasi kelas dengan Confusion Matrix

		Predicted Class		
		Yes	No	Total
Actual Class	Yes	TP	FN	P
	No	FP	TN	N
	Total	P'	N'	P+N

True Positive (TP): Jumlah data yang memiliki nilai positif secara sebenarnya dan diprediksi juga sebagai positif.
False Positive (FP): Jumlah data yang memiliki nilai negatif secara sebenarnya, tetapi diprediksi sebagai positif.
False Negative (FN): Jumlah data yang memiliki nilai positif secara sebenarnya, tetapi diprediksi sebagai negatif.
True Negative (TN): Jumlah data yang memiliki nilai negatif secara sebenarnya dan diprediksi juga sebagai negatif.

c. Kappa
   Statistik Kappa Cohen mengukur tingkat kesepakatan antara dua penilai dalam mengklasifikasikan subjek dalam salah satu dari dua kategori yang mungkin. Nilai Kappa yang mendekati satu menandakan kesepakatan yang sangat baik, sedangkan nilai mendekati nol menunjukkan kesepakatan yang lemah antara kedua penilai.
d. Precision
   Precision adalah ukuran yang mengevaluasi proporsi dokumen yang dapat ditemukan kembali oleh suatu proses pencarian dan dianggap relevan untuk kebutuhan pencarian informasi. Ini dihitung sebagai rasio jumlah dokumen relevan yang ditemukan dibagi dengan total jumlah dokumen yang ditemukan. Rumus precision dapat ditemukan untuk mengukur tingkat ketepatan hasil prediksi.

$$Precision = \frac{\text{Jumlah dokumen relevan yang terpanggil (a)}}{\text{Jumlah dokumen relevan yang ada di dalam database (a + b)}}$$

Gambar 1 Rumus Precision Algoritma Random Forest

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

e. Recall

Recall adalah proporsi jumlah dokumen yang dapat ditemukan kembali oleh sebuah proses pencarian informasi. Berikut adalah rumus untuk mnghitung angka recall.

$$Recall = \frac{\text{Jumlah dokumen relevan yang terpanggil (a)}}{\text{Jumlah dokumen relevan yang ada di dalam database (a + c)}}$$

Gambar 2 Rumus Recall Algoritma Random Forest

2.2.2 *Algoritma Support Vector Machine*

Support Vector Machine (SVM) merupakan suatu teknik yang muncul belakangan ini dan sangat diminati dalam dekade terakhir, dengan tujuan utama untuk melakukan prediksi baik pada kasus klasifikasi maupun regresi. SVM berusaha menemukan fungsi pemisah atau klasifier terbaik di antara sejumlah besar fungsi yang mungkin, dengan fokus pada pemisahan dua jenis objek. Pemisah ini, yang dikenal sebagai *hyperplane*, merupakan suatu dimensi atau struktur yang diharapkan terletak di tengah-tengah antara dua objek dari dua kelas yang berbeda.

Dalam pencarian *Hyperplane* terbaik, SVM berusaha memaksimalkan margin, yang merupakan jarak tegak lurus antara *hyperplane* dan objek terdekat yang disebut sebagai s*upport vector*. Proses optimasi dalam SVM yang bertujuan memaksimalkan nilai margin dapat dilakukan dengan strategi meminimalkan pembagi, yang dapat dijelaskan sebagai berikut.

$$minimize \frac{1}{2}||w||^2$$

$$dengan\ syarat:$$

$$y_i(wx_i + b) \geq 1, i = 1,2,\ldots,i$$

w = bobot vektor
b = bilangan skalar yang menyatakan nilai bias
w_i = data ke-i, i = 1,2,…,n
x_i = data ke-i, i = 1,2,…,n

Gambar 3 Rumus proses optimasi Algoritma Random Forest

2.2.3 *Algoritma Decision Tree C4.5*

Algoritma C4.5 menjadi salah satu algoritma yang terintegrasi dalam struktur *decision tree*. Algoritma ini adalah hasil pengembangan dari algoritma ID3 (*Iterative Dichotomiser* 3) yang awalnya dikembangkan oleh J.Ross Quinlan. Algoritma C4.5 digunakan untuk keperluan klasifikasi, segmentasi, dan pengelompokan dengan sifat yang bersifat prediktif. Prinsip dasar dari algoritma ini adalah membentuk pohon keputusan, di mana cabang-cabang dalam pohon tersebut berperan sebagai pertanyaan klasifikasi, sementara daun-daunnya merepresentasikan kelas-kelas atau segmen-segmen tertentu.

Tahapan yang dilibatkan dalam Algoritma C4.5 dapat diuraikan sebagai berikut:

1. Pemilihan atribut atau fitur untuk langkah awal.
2. Perhitungan information gain terbaik dengan menggunakan rumus:

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

Gambar 4 Rumus Algoritma C4.5 (1)

Yang mana Entropy (S) adalah nilai entropy dari S.

$$Entropy(S) = \sum_i -P_i \log P_i$$

Gambar 5 Rumus Algoritma C4.5 (2)

3. Prediksi jenis kelas untuk seluruh data pelatihan S menggunakan persamaan Entropy.

$$Entropy(S) = -P_{\oplus} \log P_{\oplus} - P_{\ominus} \log P_{\ominus}$$

Gambar 6 Rumus Algoritma C4.5 (3)

4. Perhitungan Gain dari semua atribut.
5. Perhitungan Entropy untuk masing-masing nilai variabel.
6. Seleksi atribut dengan Gain tertinggi.

3 HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi *Random Forest*

Dataset yang digunakan pada metode *Random Forest* sebanyak 70.000 data yang memiliki jumlah 12 atribut dan 1 class hasil prediksi penyakit kardiovaskular, dengan 70% data training dan 30% data testing. Dataset yang digunakan yaitu id, *age*, *height*, *weight*, *gender*, ap_hi, ap_lo, *cholesterol*, *gluc*, *smoke*, *alco*, dan *active*. Algoritma tersebut dibangun menggunakan *software Google* Colab. Evaluasi yang digunakan pada penelitian ini menggunakan *Confusion Matrix* dengan *Precision*, *Recall*, *Accuracy*, dan F1-Score. Hasil evaluasi menggunakan algoritma tersebut, menunjukkan hasil nilai *Accuracy* sebesar 71%, *Precision* sebesar 69%, *Recall* sebesar 75%, dan F1-Score sebesar 72%, yang dapat dilihat pada Gambar 7.

```
Validation precision:  0.6951564433445289
Validation recall:  0.7518401682439537
Validation accuracy:  0.7121428571428572
Validation F1-Score:  0.7223880597014926
```

Gambar 7 Hasil evaluasi Random Forest

3.2 Hasil Evaluasi *Support Vector Machine*

Dataset yang digunakan pada metode *Support Vector* Machine sebanyak 70.000 data yang memiliki jumlah 12 atribut dan 1 class hasil prediksi penyakit kardiovaskular, dengan 70% data training dan 30% data testing. Dataset yang digunakan yaitu id, *age*, *height*, *weight*, *gender*, ap_hi, ap_lo, *cholesterol*, *gluc*, *smoke*, *alco*, dan *active*. Algoritma tersebut dibangun menggunakan *software Google* Colab. Klasifikasi menggunakan *Support Vector Machine* dilakukan dengan 2 kernel yaitu RBF dan *Polynomial*. Pengujian dilakukan sebanyak 5 kali menggunakan 3 kernel RBF dengan C = default, C = 100, dan C = 0,1 serta 2 kernel Polynomial dengan C = 2 dan C = 5. Evaluasi yang digunakan pada penelitian ini menggunakan *Confusion Matrix* dengan *Accuracy*, *Precision*, *Recall*, dan F1- Score.

Hasil evaluasi pada Kernel RBF dengan C = default, menunjukkan hasil Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72% yang dapat dilihat pada Gambar 8.

```
              precision    recall  f1-score   support

           0       0.69      0.78      0.73     10352
           1       0.75      0.66      0.70     10648

    accuracy                           0.72     21000
   macro avg       0.72      0.72      0.72     21000
weighted avg       0.72      0.72      0.72     21000

array([[8077, 2275],
       [3646, 7002]])
```

Gambar 8 Hasil evaluasi kernel RBF dengan C=default

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Hasil evaluasi pada Kernel RBF dengan C = 100, menunjukkan hasil Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72% yang dapat dilihat pada Gambar 9.

	precision	recall	f1-score	support
0	0.69	0.78	0.73	10352
1	0.75	0.66	0.70	10648
accuracy			0.72	21000
macro avg	0.72	0.72	0.72	21000
weighted avg	0.72	0.72	0.72	21000

Gambar 9 Hasil evaluasi kernel RBF dengan C=100

Hasil evaluasi pada Kernel RBF dengan C = 0.1, menunjukkan hasil Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72% yang dapat dilihat pada Gambar 10.

	precision	recall	f1-score	support
0	0.69	0.78	0.73	10352
1	0.75	0.66	0.70	10648
accuracy			0.72	21000
macro avg	0.72	0.72	0.72	21000
weighted avg	0.72	0.72	0.72	21000

Gambar 10 Hasil evaluasi kernel RBF dengan C = 0.1

Hasil evaluasi pada Kernel RBF dengan C = 2, menunjukkan hasil Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72% yang dapat dilihat pada Gambar 11.

	precision	recall	f1-score	support
0	0.69	0.78	0.73	10352
1	0.75	0.66	0.70	10648
accuracy			0.72	21000
macro avg	0.72	0.72	0.72	21000
weighted avg	0.72	0.72	0.72	21000

Gambar 11 Hasil evaluasi kernel RBF dengan C = 2

Hasil evaluasi pada Kernel RBF dengan C = 5, menunjukkan hasil Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72% yang dapat dilihat pada Gambar 12.

	precision	recall	f1-score	support
0	0.69	0.78	0.73	10352
1	0.75	0.66	0.70	10648
accuracy			0.72	21000
macro avg	0.72	0.72	0.72	21000
weighted avg	0.72	0.72	0.72	21000

Gambar 12 Hasil evaluasi kernel RBF dengan C = 5

Berdasarkan hasil evaluasi, nilai pada seluruh Kernel RBF yaitu dengan C=default, C=100, C=0.1, C=2, dan C=5 mendapatkan hasil yang sama dengan nilai Accuracy sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72%.

3.3 Hasil Evaluasi *Decision Tree*

Dataset yang digunakan pada metode *Decision Tree* sebanyak 70.000 data yang memiliki jumlah 12 atribut dan 1 class hasil prediksi penyakit kardiovaskular, dengan 70% data training dan 30% data testing. Dataset yang digunakan yaitu id, age, height, weight, gender, ap_hi, ap_lo, cholesterol, gluc, smoke, alco, dan active. Algoritma tersebut dibangun menggunakan *software Google Colab*. Evaluasi yang digunakan pada penelitian ini menggunakan *Confusion Matrix* dengan *Precision*, *Recall*, *Accuracy*, dan F1-Score. Hasil evaluasi menggunakan algoritma tersebut, menunjukkan hasil nilai Accuracy sebesar 70%, Precision sebesar 70%, Recall sebesar 70%, dan F1-Score sebesar 69%, yang dapat dilihat pada Gambar 13.

	precision	recall	f1-score	support
0	0.68	0.73	0.70	11506
1	0.71	0.66	0.69	11594
accuracy			0.70	23100
macro avg	0.70	0.70	0.69	23100
weighted avg	0.70	0.70	0.69	23100

Gambar 13 Hasil evaluasi Decision Tree

4 KESIMPULAN

Berdasarkan hasil dan pembahasan pada *Cardio Vascular Diseases Dataset* yang bersumber dari Kaggle (https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset) dengan dataset berjumlah 70.000 data yang terdiri dari 13 variabel yaitu id, usia, tinggi, berat badan, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, kolesterol, glukosa, merokok, asupan alkohol, aktivitas fisik, dan kardio (ada tidaknya Penyakit Kardiovaskular), menunjukkan hasil evaluasi yang dihasilkan dari ketiga metode diantaranya metode *Random Forest*, metode *Support Vector Machine*, dan metode *Decision Tree*, didapatkan nilai hasil evaluasi tertinggi pada metode *Support Vector Machine* dengan nilai *Accuracy* sebesar 72%, Precision sebesar 72%, Recall sebesar 72%, dan F1-Score sebesar 72%, sedangkan nilai hasil evaluasi terendah pada metode *Decision Tree* dengan nilai Accuracy sebesar 70%, Precision sebesar 70%, Recall sebesar 70%, dan F1-Score sebesar 69%. Maka dari itu, dapat disimpulkan bahwa algoritma yang paling tepat untuk dataset tersebut ialah algoritma metode Random Forest.

DAFTAR PUSTAKA

[1] Sumarlinda, S., & Lestari, W. (2022, June). Aplikasi K-Nearest Neighbor (KNN) untuk Klasifikasi Penyakit Kardiovaskuler. In *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis* (pp. 259-261).

[2] Desiani, A., Akbar, M., Irmeilyana, I., & Amran, A. (2022). Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, *4*(2), 207-214.

[3] Anshori, M., Rikatsih, N., & Haris, M. S. (2022). PREDIKSI PASIEN DENGAN PENYAKIT KARDIOVASKULAR MENGGUNAKAN RANDOM FOREST. *TEKTRIKA-Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, dan Elektronika*, *7*(2), 58-64.

[4] Nugraha, W. (2021). Prediksi penyakit jantung cardiovascular menggunakan model algoritma klasifikasi. *Jurnal Sigmata*, *9*(2), 78-84.

[5] Pangaribuan, J. J., Tedja, C., & Wibowo, S. (2019). Perbandingan Metode Algoritma C4. 5 Dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner. *Journal of Informatics Engineering Research and Technology*, *1*(1).

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[6] Arifin, A., Hendyli, J., & Herwindiati, D. E. (2021). Klasifikasi Tanaman Obat Herbal Menggunakan Metode Support Vector Machine. *Computatio: Journal of Computer Science and Information Systems*, *5*(1), 25-35.

[7] Anggraiwan, Y., & Siregar, B. (2022). KLASIFIKASI HARGA MOBIL MENGGUNAKAN METODE DECISION TREE ALGORITMA C4. 5. *Computatio: Journal of Computer Science and Information Systems*, *6*(2), 70-79.

[8] Tamba, S. P. (2022). PREDIKSI PENYAKIT GAGAL JANTUNG DENGAN MENGGUNAKAN RANDOM FOREST. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, *5*(2), 176-181.

[9] Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotheraphy. *Jurnal Teknologi dan Manajemen Informatika*, *7*(1), 46-50.

[10] Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Informatik: Jurnal Ilmu Komputer*, *18*(3), 239-248.

[11] Khasanah, N., Rachman Komarudin, N. A., Maulana, Y. I., & Salim, A. (2021). Klasifikasi Kanker Kulit Menggunakan Algoritma Random Forest Skin Cancer Classification Using Random Forest Algorithm.

[12] Septiani, W. D. (2014). Penerapan Algoritma C4. 5 untuk prediksi penyakit Hepatitis. *Jurnal Techno Nusa Mandiri*, *11*(1), 69-78.

[13] Desiani, A., Akbar, M., Irmeilyana, I., & Amran, A. (2022). Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, *4*(2), 207-214.

[14] Juslim, R. R., & Herawati, F. (2018). Penyakit Kardiovaskular.

[15] Hasanah, H., & Nurmalitasari, N. (2023, January). Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung. In *STAINS (SEMINAR NASIONAL TEKNOLOGI & SAINS)* (Vol. 2, No. 1, pp. 13-18).

[16] Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari, V. (2018, March). Prediction of cardiovascular disease using machine learning algorithms. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* (pp. 1-7). IEEE.

[17] Sun, W., Zhang, P., Wang, Z., & Li, D. (2021). Prediction of cardiovascular diseases based on machine learning. *ASP Transactions on Internet of Things*, *1*(1), 30-35.

[18] Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., ... & Yan, J. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific reports*, *10*(1), 5245.

[19] Alalawi, H. H., & Alsuwat, M. S. (2021). Detection of cardiovascular disease using machine learning classification models. *International Journal of Engineering Research & Technology*, *10*(7), 151-7.

[20] Sharma, P., Saxena, K., & Sharma, R. (2016). Heart disease prediction system evaluation using C4. 5 rules and partial tree. In *Computational Intelligence in Data Mining—Volume 2: Proceedings of the International Conference on CIDM, 5-6 December 2015* (pp. 285-294). Springer India.

[21] Thenmozhi, K., & Deepika, P. (2014). Heart disease prediction using classification with different decision tree techniques. *International Journal of Engineering Research and General Science*, *2*(6), 6-11.