



Prediksi Penyakit Jantung Berbasis CRISP-DM Menggunakan Algoritma Random Forest

Icha Aurellia^{1*}, Wanda Pratiwi², Dwi Natasyah Putri³, Dedy Hartama⁴

^{1,2,3,4} Program Studi Sistem Infromasi, STIKOM Tunas Bangsa

Penulis Korespondensi: ^{1}aurelliaicha7@gmail.com, ²wandapратиwi711@gmail.com,

³dwinatasyahputri09@gmail.com, ⁴dedyhartama@amiktunasbangsa.ac.id

Abstract. Heart disease is one of the leading causes of death worldwide and is influenced by various risk factors, such as age, cholesterol levels, blood pressure, diabetes, and unhealthy lifestyles. The increasing prevalence of heart disease highlights the need for a prediction system that can support early detection quickly and accurately. This study aims to develop a heart disease prediction model using the Random Forest algorithm based on the Cross Industry Standard Process for Data mining (CRISP-DM) methodology. The dataset consisted of 1,000 patient records with 15 predictor variables and one target variable (heart disease). The research stages included Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and Area Under Curve (AUC). The results showed that the Random Forest model achieved an Accuracy of 99%, Precision of 100%, Recall of 97.44%, F1-Score of 98.70%, and an AUC of 1.00. Feature importance analysis revealed that age and cholesterol level were the most influential factors in predicting heart disease. The findings indicate that the CRISP-DM-based Random Forest algorithm can produce an accurate prediction model and has the potential to support early detection of heart disease.

Keywords: Heart Disease, Random Forest, CRISP-DM, Machine learning, Prediction.

Abstrak. Penyakit jantung merupakan salah satu penyebab utama kematian di dunia yang dipengaruhi oleh berbagai faktor risiko, seperti usia, kadar kolesterol, tekanan darah, diabetes, dan pola hidup tidak sehat. Tingginya angka kejadian penyakit jantung mendorong perlunya sistem prediksi yang mampu membantu proses deteksi dini secara cepat dan akurat. Penelitian ini bertujuan membangun model prediksi penyakit jantung menggunakan algoritma Random Forest dengan pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM). Dataset yang digunakan terdiri dari 1.000 data pasien dengan 15 variabel prediktor dan 1 variabel target (penyakit jantung). Tahapan penelitian meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, F1-score, dan area under curve (AUC). Hasil penelitian menunjukkan bahwa model Random Forest memperoleh accuracy sebesar 99%, precision sebesar 100%, recall sebesar 97,44%, F1-score sebesar 98,70%, dan AUC sebesar 1,00. Analisis feature importance menunjukkan bahwa variabel usia dan kadar kolesterol merupakan faktor yang paling berpengaruh dalam prediksi penyakit jantung. Hasil penelitian menunjukkan bahwa algoritma Random Forest berbasis CRISP-DM mampu menghasilkan model prediksi yang akurat dan berpotensi digunakan sebagai pendukung deteksi dini penyakit jantung.

Kata kunci: Penyakit Jantung, Random Forest, CRISP-DM, Machine learning, Prediksi.

1. LATAR BELAKANG

Penyakit jantung merupakan salah satu penyebab kematian paling umum di dunia dan menjadi masalah kesehatan yang terus meningkat setiap tahunnya. Usia, kadar kolesterol, tekanan darah tinggi, diabetes, kebiasaan merokok, obesitas, dan pola hidup yang kurang sehat adalah beberapa faktor risiko yang berkontribusi terhadap tingginya

angka kejadian penyakit jantung[1]. Akibatnya, deteksi dini sangat penting untuk mencegah dan menangani penyakit dengan lebih baik.

Data mining dan *machine learning* telah berkembang dalam bidang kesehatan sebagai akibat dari kemajuan teknologi informasi dan peningkatan ketersediaan data kesehatan[2]. Dengan teknologi ini, data dapat dianalisis secara cepat dan akurat untuk membantu pengambilan keputusan medis. Ini dapat digunakan untuk mengklasifikasikan penyakit berdasarkan karakteristik pasien, sehingga pola yang terbentuk dari data historis dapat digunakan untuk memperkirakan risiko penyakit[3].

Untuk memprediksi penyakit jantung, berbagai algoritma *machine learning* telah digunakan, termasuk Random Forest, Decision Tree, Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Logistic Regression, dan Support Vector Machine. Random Forest memiliki keunggulan dalam menangani data berdimensi besar, mengurangi risiko *overfitting*, dan menghasilkan performa klasifikasi yang tinggi melalui pendekatan pembelajaran kelompok[4]. Selain itu, Random Forest dapat menghitung tingkat kepentingan setiap variabel dengan analisis *feature importance*.

Agar pengembangan model prediksi penyakit jantung dapat dilakukan secara sistematis dan mudah direplikasi, diperlukan pendekatan sistematis[5]. Ini adalah kenyataan bahwa banyak penelitian tentang topik ini telah dilakukan. Cross Industry Standard Process for *Data Mining* (CRISP-DM), yang terdiri dari tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan adalah salah satu metodologi yang paling umum digunakan dalam proyek *data mining*[6].

Dengan demikian, berdasarkan berbagai faktor risiko kesehatan pasien, penelitian ini menerapkan algoritma Random Forest dengan pendekatan CRISP-DM untuk memprediksi penyakit jantung[7]. Tujuan dari penelitian ini adalah untuk membuat model prediksi penyakit jantung dan mengevaluasi kinerjanya dengan menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Area Under Curve (AUC)*. Selain itu, faktor-faktor yang paling signifikan dalam prediksi penyakit jantung akan diidentifikasi melalui analisis *feature importance*.

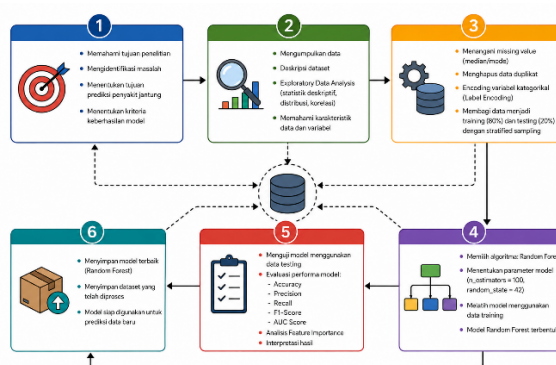
2. METODE PENELITIAN

Penelitian ini membuat model prediksi penyakit jantung dengan *data mining*. Metode Cross Industry Standard for *Data Mining* (CRISP-DM) dipilih karena menyediakan kerangka kerja yang sistematis dan terorganisir untuk pengembangan

model *machine learning* dari identifikasi masalah hingga implementasi model[8]. Dalam penelitian ini, algoritma Random Forest digunakan untuk mengkategorikan status penyakit jantung pasien berdasarkan faktor risiko kesehatan mereka.

Tahapan Penelitian CRISP-DM

Penelitian ini melakukan enam tahapan utama dari metodologi CRISP-DM, yaitu Pemahaman Bisnis (*Business Understanding*), Pemahaman Data (*Data Understanding*), Persiapan Data (*Data Preparation*), Pemodelan (*Modeling*), Evaluasi (*Evaluation*), dan Penerapan (*Deployment*). Alur penelitian yang digunakan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Menggunakan Metodologi CRISP-DM

Melalui tahapan tersebut, proses pengembangan model dilakukan secara sistematis mulai dari pemahaman permasalahan penelitian, pengolahan data, pembangunan model, hingga evaluasi dan penyimpanan model yang dihasilkan.

Pemahaman Bisnis

Tujuan dari tahap *Pemahaman Bisnis* adalah untuk memahami masalah penelitian dan menetapkan tujuan. Salah satu penyebab kematian paling umum di seluruh dunia adalah penyakit jantung, yang dipengaruhi oleh berbagai faktor risiko, termasuk usia, tekanan darah, kadar kolesterol, diabetes, obesitas, kebiasaan merokok, dan pola hidup yang kurang sehat[9]. Dengan meningkatnya jumlah kasus penyakit jantung, sistem prediksi yang dapat membantu proses deteksi dini secara lebih cepat dan akurat sangat diperlukan[10].

Penelitian ini bertujuan untuk membangun model prediksi penyakit jantung menggunakan algoritma Random Forest dengan pendekatan CRISP-DM. Model ini diharapkan dapat membantu proses identifikasi risiko penyakit jantung berdasarkan karakteristik kesehatan pasien.

Pemahaman Data

Tahap *Data Understanding* dilakukan untuk memahami karakteristik dataset yang digunakan dalam penelitian. Dataset terdiri dari 1.000 data pasien dengan 15 variabel prediktor dan 1 variabel target (penyakit jantung). Variabel target bernilai 0 untuk pasien yang tidak terindikasi penyakit jantung dan 1 untuk pasien yang terindikasi penyakit jantung. Variabel yang digunakan dalam penelitian disajikan pada Tabel 1.

Tabel 1. Deskripsi Variabel Penelitian

No	Variabel	Keterangan
1	Usia	Usia pasien
2	Jenis Kelamin	Jenis kelamin pasien
3	Kolesterol	Kadar kolesterol
4	Tekanan Darah	Tekanan darah
5	Detak Jantung	Denyut jantung
6	Merokok	Status merokok
7	Konsumsi Alkohol	Konsumsi alcohol
8	Jam Olahraga	Durasi olahraga
9	Riwayat Keluarga	Riwayat penyakit jantung keluarga
10	Diabetes	Status <i>diabetes</i>
11	Obesitas	Status obesitas
12	Tingkat Stres	Tingkat stress
13	Gula Darah	Kadar gula darah
14	Angina yang Dipicu oleh Aktivitas Fisik/Olahraga	Nyeri dada akibat aktivitas fisik
15	Jenis Nyeri Dada	Jenis nyeri dada
16	Penyakit Jantung	Status penyakit jantung (target)

Pada tahap ini dilakukan analisis statistik deskriptif dan *Exploratory Data Analysis (EDA)* untuk mengetahui distribusi data serta karakteristik setiap variabel sebelum dilakukan proses pemodelan.

Persiapan Data

Proses *Data Preparation* dilakukan untuk menghasilkan dataset yang siap untuk digunakan dalam proses pembangunan model pembelajaran mesin. Tahap persiapan termasuk memeriksa dan menangani nilai yang tidak ada, menghapus data duplikat, dan

mengubah data kategorikal menjadi data numerik dengan menggunakan teknik *Label Encoding*[11]. Untuk atribut numerik, median dan modus digunakan untuk menangani nilai yang hilang. Untuk menghindari redundansi yang dapat mempengaruhi kualitas model, data duplikat dihapus.

Karena Random Forest adalah algoritma berbasis pohon keputusan yang tidak sensitif terhadap perbedaan skala antar atribut, tidak ada normalisasi atau standarisasi data yang dilakukan dalam penelitian ini. Setelah pembersihan dan transformasi selesai, dataset dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20 menggunakan metode *Split Train-Test*. Untuk memastikan proporsi kelas pada variabel target, *stratified sampling* digunakan[12].

Pemodelan

Tahap *Modeling* dilakukan dengan membangun model klasifikasi menggunakan algoritma Random Forest. Random Forest merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan (*Decision Tree*) dan menggabungkan hasil prediksi dari seluruh pohon untuk menghasilkan keputusan akhir yang lebih stabil dan akurat[13].

Prediksi akhir Random Forest diperoleh menggunakan mekanisme majority voting yang dapat dirumuskan sebagai berikut:

$$\hat{Y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x))$$

Keterangan:

$\hat{Y}$ = hasil prediksi akhir

$h_1(x), h_2(x), \dots, h_n(x)$ = hasil prediksi setiap pohon keputusan

n = jumlah pohon keputusan

mode = nilai yang paling sering muncul

Pada penelitian ini model Random Forest dibangun menggunakan parameter $n_estimators = 100$ dan $random_state = 42$. Proses pelatihan model dilakukan menggunakan data training yang telah melalui tahap *Data Preparation*.

Evaluasi

Tahap *Evaluation* dilakukan untuk mengevaluasi kinerja model Random Forest. Ini dilakukan dengan menggunakan data pengujian, atau testing data, yang tidak digunakan selama proses pelatihan. Tujuan dari evaluasi adalah untuk mengetahui kemampuan model untuk secara akurat mengklasifikasikan status penyakit jantung[7].

Pada penelitian ini, evaluasi dilakukan menggunakan *Confusion Matrix* yang terdiri dari *True Positive (TP)*, *True Negative (TN)*, *False positive (FP)*, dan *False Negative (FN)*.

Berdasarkan *Confusion Matrix*, performa model diukur menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC)[14]. Accuracy mengukur tingkat ketepatan klasifikasi secara keseluruhan, Precision mengukur ketepatan prediksi positif, Recall mengukur kemampuan model untuk menemukan kasus positif, dan F1-Score mengukur keseimbangan antara Precision dan Recall. Selain itu, AUC mengukur kemampuan model untuk membedakan antara kasus positif dan negatif[15]. Hasil evaluasi ini digunakan untuk menentukan tingkat keandalan model untuk memprediksi penyakit jantung.

Accuracy

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam melakukan klasifikasi data.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Keterangan:

TP = True Positive

TN = True Negative

FP = *False positive*

FN = False Negative

Precision

Precision digunakan untuk mengukur proporsi prediksi positif yang benar terhadap seluruh prediksi positif yang dihasilkan model.

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

Keterangan:

TP = True Positive

FP = *False positive*

Recall

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data positif yang sebenarnya.

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

Keterangan:

TP = True Positive

FN = False Negative

F1-Score

F1-Score merupakan rata-rata harmonik antara Precision dan Recall sehingga memberikan ukuran performa yang lebih seimbang.

$$F1-Score = 2 \frac{(Precision \times Recall)}{(Precision + Recall)}$$

Keterangan:

Precision = nilai precision

Recall = nilai recall

Area Under Curve (AUC)

Area Under Curve (AUC) digunakan untuk mengukur kemampuan model dalam membedakan kelas positif dan negatif pada berbagai nilai *threshold*[7]. Kemampuan model untuk melakukan klasifikasi menjadi lebih baik semakin mendekati nilai 1

Selain menilai kinerja model, penelitian ini juga menganalisis pentingnya fitur untuk mengetahui fitur mana yang paling banyak berkontribusi pada prediksi penyakit jantung.

Penerapan

Tahap *Deployment* adalah tahap terakhir dari metodologi CRISP-DM. Di sini, model Forest Random yang telah dilatih disimpan dalam bentuk file dan dapat digunakan kembali untuk mengantisipasi data baru[16]. Selain itu, dataset yang telah melalui proses pembersihan dan transformasi juga disimpan sebagai dataset akhir untuk analisis lebih lanjut dan pengembangan sistem prediksi penyakit jantung berbasis *machine learning*.

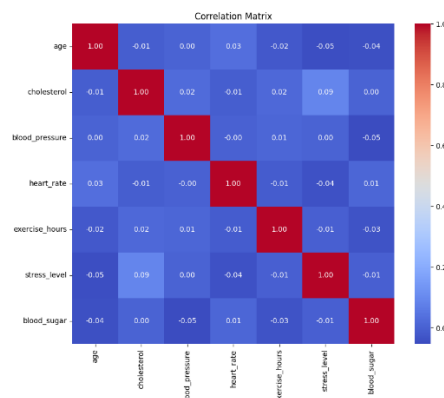
3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, algoritma Random Forest digunakan untuk membangun model prediksi penyakit jantung. Pendekatan Cross Industry Standard Process for *Data Mining* (CRISP-DM) juga digunakan. Penelitian telah dilakukan secara sistematis mulai

dari Pemahaman Bisnis (*Business Understanding*), Pemahaman Data (*Data Understanding*), Persiapan Data (*Data Preparation*), Pemodelan (*Modeling*), Evaluasi (*Evaluation*), dan Penerapan (*Deployment*). Hasil penelitian meliputi hasil *Exploratory Data Analysis* (EDA), hasil evaluasi model Random Forest, dan analisis tingkat kepentingan variabel (*feature importance*) yang mempengaruhi prediksi penyakit jantung.

Hasil Exploratory Data Analysis (EDA)

Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk mempelajari karakteristik dataset. Digunakan 1.000 data pasien dengan 15 variabel prediktor dan 1 variabel target, penyakit jantung. Untuk mengetahui bagaimana variabel yang digunakan dalam penelitian berhubungan satu sama lain, analisis dilakukan dengan visualisasi data dan statistik deskriptif.



Gambar 2. Korelasi Antar Variabel Numerik

Berdasarkan Gambar 2, beberapa variabel numerik menunjukkan hubungan terhadap status penyakit jantung, terutama usia (*age*), kadar kolesterol (*cholesterol*), tekanan darah (*blood_pressure*), kadar gula darah (*blood_sugar*), dan denyut jantung (*heart_rate*). Variabel-variabel tersebut merupakan faktor kesehatan yang secara teoritis berperan dalam meningkatkan risiko penyakit jantung sehingga berpotensi memengaruhi hasil prediksi model.

Tahap EDA juga mencakup pemeriksaan kualitas data melalui identifikasi nilai yang hilang dan data duplikat. Nilai yang hilang ditangani dengan metode imputasi yang sesuai, dan data duplikat dihapus untuk menjaga kualitas dataset. Hasil proses

menunjukkan bahwa dataset memiliki fitur yang bagus dan siap digunakan selama proses pemodelan. Tidak ada masalah signifikan yang mengganggu kinerja model.

Hasil Evaluasi Model Random Forest

Setelah melalui tahap *Data Preparation*, dataset dibagi menjadi data pelatihan (*training data*) sebesar 80% dan data pengujian (*testing data*) sebesar 20%. Model Random Forest kemudian dibangun menggunakan parameter *n_estimators* sebanyak 100 pohon keputusan dan *random_state* sebesar 42.

Evaluasi model dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC). Hasil evaluasi model Random Forest disajikan pada Tabel 2.

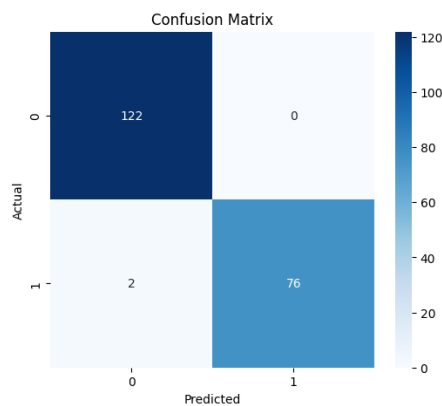
Tabel 2. Hasil Evaluasi Model Random Forest

Metrik Evaluasi	Nilai
Accuracy	0,99
Precision	1,00
Recall	0,9744
F1-Score	0,9870
AUC Score	1,00

Berdasarkan Tabel 2, model Random Forest memperoleh nilai Accuracy sebesar 99%, Precision sebesar 100%, Recall sebesar 97,44%, F1-Score sebesar 98,70%, dan AUC sebesar 1,00. Hasil menunjukkan bahwa model dapat mengklasifikasikan status penyakit jantung dengan sangat baik dan sangat akurat. Mereka juga sangat baik dalam mengurangi kesalahan prediksi positif, atau *false positive*.

Nilai Recall yang mencapai 97,44% dan nilai F1-Score yang tinggi menunjukkan keseimbangan yang baik antara Precision dan Recall, sedangkan nilai AUC sebesar 1,00 menunjukkan bahwa sebagian besar kasus penyakit jantung diidentifikasi dengan benar. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma Random Forest dapat memprediksi penyakit jantung dengan sangat baik pada dataset yang digunakan.

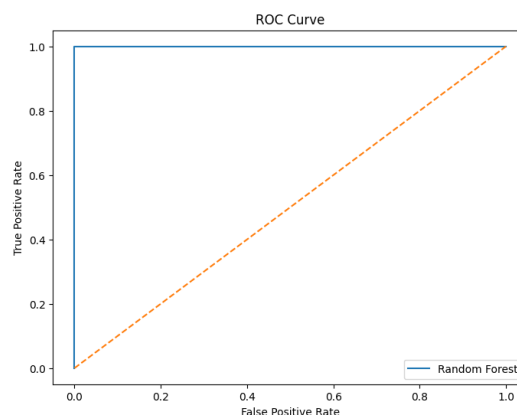
Untuk melihat performa klasifikasi secara lebih rinci, dilakukan analisis menggunakan *Confusion Matrix*.



Gambar 3. *Confusion Matrix* Random Forest

Gambar 3 menunjukkan bahwa model memiliki tingkat kesalahan klasifikasi yang sangat rendah, seperti yang ditunjukkan oleh jumlah prediksi yang sesuai antara kelas aktual dan kelas hasil prediksi. Hasil ini mendukung temuan pada metrik evaluasi, yang menunjukkan bahwa model memiliki kemampuan untuk melakukan klasifikasi dengan sangat baik.

Selain *Confusion Matrix*, evaluasi model juga dilakukan menggunakan kurva *Receiver Operating Characteristic (ROC)*.

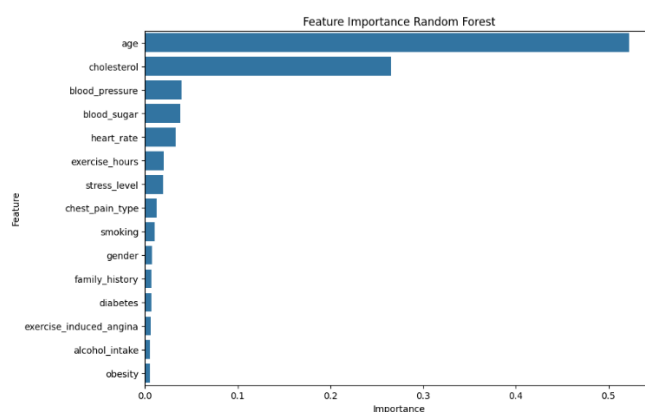


Gambar 4. Kurva ROC Model Random Forest

Kurva ROC sangat dekat dengan sudut kiri atas grafik, seperti yang ditunjukkan pada Gambar 4. Model memiliki kemampuan yang sangat baik untuk membedakan pasien yang terindikasi penyakit jantung dari mereka yang tidak. Posisi kurva menunjukkan ini. Nilai AUC sebesar 1,00 menunjukkan bahwa model Random Forest dapat melakukan klasifikasi data dengan sangat akurat.

Analisis Feature importance dan Pembahasan

Algoritma Random Forest tidak hanya memiliki kemampuan untuk menghasilkan hasil klasifikasi yang sangat baik, tetapi juga mampu mengidentifikasi tingkat kepentingan masing-masing variabel melalui analisis *feature importance*. Analisis ini digunakan untuk menentukan komponen yang paling berpengaruh terhadap prediksi penyakit jantung.



Gambar 5. *Feature importance* pada Model Random Forest

Variabel usia (*age*) memiliki tingkat signifikansi tertinggi, dengan nilai signifikansi 0,522561, diikuti oleh kadar kolesterol (*cholesterol*) dengan nilai signifikansi 0,265493, seperti yang ditunjukkan pada Gambar 5. Hasil ini menunjukkan bahwa dua faktor utama yang memengaruhi prediksi penyakit jantung adalah usia dan kadar kolesterol. Tekanan darah (*blood pressure*), gula darah (*blood sugar*), dan denyut jantung (*heart rate*) juga berkontribusi, tetapi dengan tingkat pengaruh yang lebih rendah.

Meskipun demikian, faktor-faktor seperti jumlah waktu yang dihabiskan untuk berolahraga, tingkat stres, jenis nyeri dada, jenis kelamin, riwayat keluarga, *diabetes* exercise induced angina, konsumsi alkohol, dan obesitas memiliki dampak yang sangat kecil. Sebaliknya, variabel-variabel tersebut masih membantu model menemukan pola yang terkait dengan penyakit jantung.

Secara keseluruhan, hasil analisis *feature importance* menunjukkan bahwa usia dan kadar kolesterol adalah sinyal paling penting untuk memprediksi penyakit jantung. Hasil ini menunjukkan bahwa algoritma Random Forest tidak hanya mampu menghasilkan prediksi yang sangat baik dengan nilai akurasi 99 persen, ketepatan 100 persen, pengembalian 97,44 persen, skor F1 98,70 persen, dan AUC 1,00, tetapi juga mampu menjelaskan komponen yang paling berpengaruh terhadap risiko penyakit jantung.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa algoritma Random Forest dan metodologi CRISP-DM berhasil digunakan untuk membangun model prediksi penyakit jantung berdasarkan berbagai faktor risiko kesehatan pasien. Hasil evaluasi menunjukkan bahwa model ini memiliki performa yang sangat baik dengan nilai akurasi sebesar 99%, ketepatan sebesar 100%, recall sebesar 97,44%, skor F1-sebesar 98,70%, dan AUC sebesar 1,00. Ini menunjukkan bahwa algoritma Random Forest dapat mengklasifikasikan status penyakit dengan benar.

Menurut analisis karakteristik penting, variabel usia (*age*) dan kadar kolesterol (*cholesterol*) adalah yang paling berpengaruh dalam menentukan kemungkinan penyakit jantung. Oleh karena itu, penelitian ini menunjukkan bahwa algoritma Random Forest berbasis CRISP-DM dapat digunakan sebagai dasar untuk membuat sistem pendukung keputusan yang membantu mendeteksi penyakit jantung secara lebih cepat dan akurat.

DAFTAR REFERENSI

- [1] R. Diah Ruswanti, Dahlan Susilo, "Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45," *GO INFOTECH J. Ilm. STMIK AUB*, vol. 30, no. 1, pp. 111–121, 2024, doi: 10.36309/goi.v30i1.266.
- [2] F. S. Rakha Gusti Wardhana, Gunawan Wang, "PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA," *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023, doi: <https://doi.org/10.24076/joism.2023v5i1.1136>.
- [3] M. A. Faradeya and E. R. Subhiyakto, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes," *J. Algoritm.*, vol. 11, no. 1, pp. 115–127, 2025, doi: 10.33364/algoritma/v.22-1.2178.
- [4] D. I. J. Ihsanuddin Akbar, Fidi Supriadi, "PEMANFAATAN MACHINE LEARNING DI BIDANG KESEHATAN," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1744–1749, 2025, doi: <https://doi.org/10.36040/jati.v9i1.12663>.
- [5] R. Julkarnaen Karo-Karo, Adryan Raihan Syakir, Raihan, Sumanto, Imam Budiawan and A. C. Pakpahan, "Heart Disease Prediction Using Machine Learning Algorithms Based on Health Indicators with Random Forest," *J. Ilm. Sist. Inf.*, vol. 5, no. 1, pp. 84–95, 2026, doi: <https://doi.org/10.51903/ltgz3234>.
- [6] M. F. R. M. Arya Ardhi Baskara, Nurul Maharani Piranti, "FRAMEWORK DATA MINING : SEBUAH SURVEI," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4886–4895, 2025, doi: <https://doi.org/10.36040/jati.v9i3.13803>.
- [7] A. S. B. Maulidina Cahaya Rani, Farah Diba Azkia, Revinta Arrova Dewi, Mochamad Wahyudi, Sumanto, "PERBANDINGAN ALGORITMA RANDOM FOREST, NAIVE BAYES, DAN NEURAL NETWORK DALAM KLASIFIKASI PENYAKIT JANTUNG," *J. Sains Inform. Terap.*, vol. 4, no. 2, pp. 77–84, 2025, doi: <https://doi.org/10.62357/jsit.v4i2.609>.
- [8] A. P. B. K. Muhammad Fauzani Akbar, Ita Damaris Br Munthe, Albert Julian Marito Silaban, Ahmad Dwi Abiyyu, "Literature Reveiw Penerapan Machine

- Learning Pada Supply Chain Management,” *J. Sist. Informasi, Tek. Komput. dan Teknol. Pendidik.*, vol. 5, no. M1, pp. 302–310, 2026, doi: <https://doi.org/10.55338/justikpen.v5i2.427>.
- [9] N. D. S. Irfan Sazali Nasution, Arini Dwi Rahmadani, Windi Audina, Dinda Purnama Sari, “Systematic Review: Pengaruh Gaya Hidup dan Pengetahuan Masyarakat terhadap Risiko Penyakit Jantung Koroner,” *SEHATRAKYAT (Jurnal Kesehat. Masyarakat)*, vol. 4, no. 2, pp. 287–298, 2025, doi: [10.54259/sehatrakyat.v4i2.4337](https://doi.org/10.54259/sehatrakyat.v4i2.4337).
- [10] K. C. J. S. Prasetyo2, “Prediksi Penyakit Jantung Koroner Menggunakan Metode K-NN dan Regresi Logistik Berdasarkan Kerangka Kerja CRISP-DM,” *Semin. Nas. Univ. MA CHUNG 2024*, vol. 4, no. 2020, pp. 241–248, 2024.
- [11] I. Y. Maulana and B. Irawan, “Rancangan Sistem Prediksi Penyakit Jantung Berbasis Framingham Risk Score: Konsistensi Teoritis dan Implementasi Web,” *JCBD J. Comput. Digit. Bus.*, vol. 4, no. 3, pp. 146–151, 2025, doi: <https://doi.org/10.56427/jcbd.v4i3.775>.
- [12] A. Y. Yoga Adi Prasetyo, Ema Utami, “Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM,” *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 382–390, 2024, doi: [10.33650/jeecom.v4i2](https://doi.org/10.33650/jeecom.v4i2).
- [13] A. S. R. Ergi Putra Febtiawan, Lalu A. Syamsul Irfan Akbar, “Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification,” *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 1053–1062, 2024, doi: [10.47065/josh.v5i4.5514](https://doi.org/10.47065/josh.v5i4.5514).
- [14] N. S. Adinda Dwi Putri, Fitriana Sholekhah, Eric Dadynata, Lusiana Efrizoni, Rahmadden, “The Application of C4 . 5 Decision Tree Algorithm for Predicting the Survival Rate of Thyroid Cancer Patients Penerapan Algoritma Decesion Tree C4 . 5 untuk Memprediksi Tingkat Kelangsungan Hidup Pasien Kanker Tiroid,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. October, pp. 1485–1495, 2024, doi: <https://doi.org/10.57152/malcom.v4i4.1532>.
- [15] M. Muhammad, J. Samodro, E. M. Sakti, and B. M. Kumudasmoro, “Analisis Pengaruh Ketidakseimbangan Data terhadap Kinerja Model Klasifikasi Penyakit Jantung,” *J. Softw. Eng. Inf. Syst. e-ISSN*, vol. 6, no. 1, pp. 56–62, 2026.
- [16] D. B. Saputra *et al.*, “PENERAPAN MODEL CRISP-DM PADA PREDIKSI NASABAH KREDIT MENGGUNAKAN ALGORITMA RANDOM FOREST,” *Idealis Indones. J. Inf. Syst.*, vol. 7, no. 2021, pp. 240–247, 2024.