



## Sistem Pendeteksi Penyakit Jantung Berbasis Kecerdasan Buatan Menggunakan Metode Random Forest

Rendy Fransiskus Silaban<sup>1</sup>, Nur Hidayat Attamim<sup>2</sup>, Muhammad Juliadi<sup>3</sup>,  
Laura Cristina Natalia<sup>3</sup>

<sup>1, 2, 3, 4</sup> Universitas Satya Terra Bhinneka

Alamat: Jl. Sunggal Gg. Bakul, Sunggal, Kec. Medan Sunggal, Kota Medan, Sumatera Utara

Korespondensi penulis: [fransiskusrendy28@gmail.com](mailto:fransiskusrendy28@gmail.com)

**Abstract.** This research focuses on the development of an early detection system for heart disease using Artificial Intelligence (AI) and the Random Forest method. Heart disease, particularly coronary heart disease (CHD), is a leading cause of death worldwide, with early detection being a crucial factor in reducing mortality rates. The study uses a dataset from the UCI Machine Learning Repository, consisting of 303 patient records with various medical factors like age, sex, cholesterol levels, and blood pressure. The Random Forest algorithm is employed to create a predictive model capable of classifying whether an individual is at risk for heart disease. Data preprocessing, including normalization and encoding categorical variables, is carried out before training the model. The system achieved a high accuracy rate of 98.5% in predicting heart disease risk. The developed model is deployed as an API and can be integrated into healthcare applications for real-time risk prediction, supporting timely medical decisions. This system aims to contribute to the improvement of early heart disease detection, offering an efficient tool for healthcare professionals.

**Keywords:** Artificial Intelligence, heart disease, Random Forest

**Abstrak.** Penelitian ini berfokus pada pengembangan sistem deteksi dini penyakit jantung menggunakan Kecerdasan Buatan (AI) dan metode Random Forest. Penyakit jantung, khususnya penyakit jantung koroner (PJK), merupakan penyebab utama kematian di seluruh dunia, dengan deteksi dini menjadi faktor penting dalam mengurangi angka kematian. Penelitian ini menggunakan dataset dari UCI Machine Learning Repository yang terdiri dari 303 data pasien dengan berbagai faktor medis seperti usia, jenis kelamin, kadar kolesterol, dan tekanan darah. Algoritma Random Forest digunakan untuk membuat model prediksi yang mampu mengklasifikasikan apakah seseorang berisiko terkena penyakit jantung. Pra-pemrosesan data, termasuk normalisasi dan encoding variabel kategorikal, dilakukan sebelum melatih model. Sistem ini mencapai tingkat akurasi tinggi sebesar 98,5% dalam memprediksi risiko penyakit jantung. Model yang dikembangkan diimplementasikan sebagai API dan dapat diintegrasikan ke dalam aplikasi kesehatan untuk prediksi risiko secara real-time, mendukung pengambilan keputusan medis yang lebih tepat waktu. Sistem ini diharapkan dapat berkontribusi pada peningkatan deteksi dini penyakit jantung, menawarkan alat yang efisien bagi tenaga medis.

**Kata kunci:** Kecerdasan Buatan, penyakit jantung, Random Forest,

### 1. LATAR BELAKANG

Penyakit jantung adalah penyakit kronis yang menyebabkan kematian secara global. Menurut *World Health Organization* (WHO), sekitar 17,9 juta orang meninggal setiap tahun akibat penyakit kardiovaskular, yang mencakup serangan jantung dan stroke (Haldy & Meily Kurniawidjaja, 2024). Kondisi ini diperparah oleh kurangnya deteksi dini, gaya hidup tidak sehat, dan kurangnya akses terhadap layanan kesehatan yang cepat dan tepat. Oleh karena itu, pengembangan sistem deteksi dini penyakit jantung menjadi sangat krusial. Penyakit jantung koroner (PJK) sering kali tidak terdeteksi pada tahap awal, karena gejalanya cenderung muncul hanya saat kondisi pasien sudah berada pada tahap yang lebih parah. Hal ini memperburuk angka kematian yang dapat dicegah dengan deteksi lebih awal (Wijaya et al., 2025). Teknologi

yang mampu mendeteksi penyakit jantung lebih cepat dan lebih akurat akan sangat membantu dalam mengurangi angka kematian akibat penyakit ini.

Dengan perkembangan teknologi yang pesat, AI menjadi solusi efektif dan efisien untuk mendeteksi penyakit secara otomatis. Salah satu metode AI yang telah terbukti efektif dalam klasifikasi data medis adalah algoritma *Random Forest* (Elektronik et al., 2023). Algoritma ini mampu mengelola data dengan banyak variabel, memiliki tingkat akurasi tinggi, dan relatif tahan terhadap *overfitting*. Random Forest berproses dengan membentuk beberapa pohon keputusan yang kemudian digabungkan untuk meningkatkan akurasi prediksi, sehingga dapat memberi hasil yang sesuai dalam mendeteksi risiko penyakit jantung (Alfajr & Defiyanti, 2024). Keunggulan algoritma ini terletak pada kemampuannya untuk menangani data yang bervariasi dan tidak terstruktur dengan baik, yang sering ditemukan dalam data medis.

Dengan memanfaatkan dataset penyakit jantung dari *UCI Machine Learning Repository*, penelitian ini merancang sistem prediksi berbasis algoritma *Random Forest* yang diimplementasikan menggunakan *platform Node.js* dan disediakan dalam bentuk layanan API. Sistem ini diharapkan dapat membantu dalam mendeteksi risiko penyakit jantung secara real-time dan akurat, serta dapat diintegrasikan ke berbagai aplikasi kesehatan (Nurrohman & Kurniawan, 2025). Keuntungan menggunakan *Node.js* adalah kemampuannya untuk menangani banyak permintaan secara bersamaan, memungkinkan sistem bekerja secara efisien dalam skala besar. Selain itu, dengan menyediakan layanan API, sistem ini dapat diintegrasikan dengan mudah ke dalam aplikasi kesehatan lainnya, seperti aplikasi pemantauan kesehatan berbasis ponsel pintar, yang dapat memberikan informasi penting kepada pengguna atau tenaga medis secara cepat dan akurat. Sistem ini diharapkan menjadi alat bantu yang dapat mendukung pengambilan keputusan medis yang lebih baik, serta meningkatkan kualitas pelayanan kesehatan dalam deteksi dini penyakit jantung.

## 2. KAJIAN TEORITIS

### Penyakit Jantung

Penyakit jantung adalah gangguan pada jantung dimana tidak bisa melaksanakan fungsinya secara optimal sebagai pemompa darah bagi tubuh, sehingga terjadi gangguan pada sistem peredaran darah (Pashar et al., 2025). Gangguan ini dapat terjadi karena melemahnya otot jantung, adanya kelainan struktural seperti celah pada serambi jantung, atau gangguan lain yang menyebabkan jantung gagal memompa darah secara efektif (Wijaya et al., 2025). Penyakit ini dapat terjadi pada semua kalangan tanpa memandang aspek tertentu dan umumnya bersifat kronis serta sulit untuk disembuhkan sepenuhnya.

Penyakit jantung tidak hanya menyerang kelompok usia lanjut, namun juga dapat terjadi pada usia produktif. Penyakit ini sering kali berkembang tanpa gejala yang jelas hingga terjadi komplikasi berat seperti serangan jantung atau gagal jantung. Selain itu, penyakit jantung juga mencakup kondisi gagal jantung, yaitu berkurangnya kinerja jantung dalam memompa darah secara efektif (Rahmawati et al., 2025).

Penyakit jantung koroner (PJK) merupakan salah satu jenis yang paling rentan dan menjadi penyebab utama kematian di global, termasuk di Indonesia. PJK terjadi akibat penyumbatan pembuluh darah koroner yang memasok darah ke otot jantung, umumnya disebabkan oleh penumpukan plak kolesterol di dinding arteri (aterosklerosis) (Santosa & Baharuddin, 2020).

### **Kecerdasan Buatan (*Artificial Intelligence/AI*)**

*Artificial Intelligence* (AI) adalah cabang ilmu komputer yang berfokus pada system development atau machine yang mampu meniru kecerdasan manusia, seperti belajar, bernalar, memecahkan masalah, memahami bahasa, dan membuat keputusan. AI bekerja dengan memanfaatkan data, algoritma, dan kekuatan komputasi untuk menganalisis informasi, mengenali pola, serta membuat prediksi atau keputusan secara otomatis (Afandi & Kurnia, 2023). *Artificial Intelligence* (AI) secara umum didefinisikan sebagai teknologi yang memungkinkan komputer dan mesin untuk mengimitasi kemampuan manusia dalam belajar, memahami, memecahkan masalah, mengambil keputusan, berkreasi, dan bertindak secara otonom. Sistem dan perangkat yang dilengkapi AI dapat mengenali objek, memahami serta merespons bahasa manusia, belajar dari pengalaman, serta membuat rekomendasi atau keputusan secara mandiri tanpa intervensi manusia langsung (Wahyudi, 2023).

AI dapat dikategorikan menjadi dua tipe utama, yaitu *Narrow AI* yang dirancang untuk melakukan tugas khusus seperti deteksi wajah atau suara, dan *General AI* yang bertujuan meniru kemampuan kognitif manusia secara menyeluruh, meski hingga kini masih bersifat teoretis (Sobron & Lubis, 2021). AI saat ini banyak diaplikasikan dalam berbagai bidang, mulai dari pengenalan suara, pengolahan citra, hingga kendaraan otonom dan asisten virtual. Dalam perkembangannya, AI tidak hanya mampu menjalankan tugas rutin, tetapi juga mampu menangani tantangan kompleks seperti menganalisis data medis dalam jumlah besar atau mendeteksi pola pada citra medis yang hanya bisa diproses oleh manusia (Nuryani et al., 2023).

Dalam dunia medis, AI telah menjadi teknologi yang sangat transformatif. AI di bidang kesehatan umumnya diterapkan dalam bentuk *machine learning* dan *deep learning* untuk menganalisa data medis, mendukung pengambilan keputusan klinis, serta meningkatkan efisiensi dan akurasi diagnosis (Sanhaji & Hizbullah, 2023). Contoh penerapan AI di dunia medis meliputi analisis citra radiologi (*CT scan, MRI, X-ray*), pengembangan obat, prediksi

risiko penyakit, personalisasi terapi, hingga otomatisasi tugas administratif seperti penjadwalan janji temu dan pengelolaan rekam medis (Atmojo et al., 2024). AI juga dapat membantu dokter dalam mendeteksi penyakit lebih awal, memberikan rekomendasi pengobatan yang dipersonalisasi, serta mengurangi beban kerja administratif sehingga dokter bisa lebih fokus pada pelayanan pasien. Selain itu, AI berperan penting dalam penelitian dan pengembangan obat baru, mempercepat proses yang biasanya memakan waktu bertahun-tahun menjadi lebih singkat dan efisien (Trenggono & Bachtiar, 2023). Dengan kemampuannya menganalisis data dalam skala besar, AI membantu meningkatkan kualitas layanan kesehatan, menurunkan biaya, dan memperluas akses terhadap pelayanan medis yang berkualitas.

### **Random Forest**

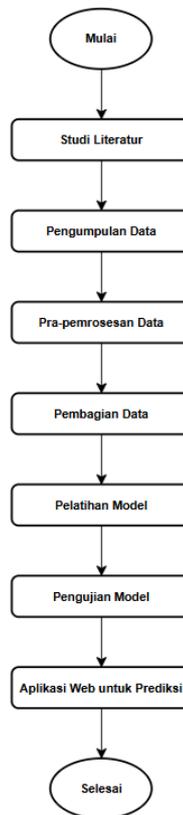
Random Forest adalah salah satu algoritma machine learning berbasis *ensemble* yang sangat populer dan sering digunakan karena kemampuannya dalam menangani masalah klasifikasi maupun regresi secara efektif (Jalan et al., 2024). Algoritma ini pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 dan bekerja dengan membangun banyak pohon keputusan (*decision tree*) secara acak dari subset data dan fitur yang berbeda-beda. Setiap pohon keputusan dalam Random Forest akan memberikan prediksi, dan hasil akhir ditentukan berdasarkan voting mayoritas (untuk klasifikasi) atau rata-rata (untuk regresi) dari seluruh pohon yang dibangun (All et al., 2006). Pendekatan ini membuat *Random Forest* mampu mengurangi risiko *overfitting* yang sering terjadi, serta meningkatkan akurasi dan stabilitas prediksi.

*Random Forest* menggunakan dua sumber keacakan utama, yaitu pertama, dengan mengambil sampel bootstrap secara acak dari data pelatihan untuk membangun setiap pohon, dan kedua, dengan memilih subset acak dari fitur pada setiap *node split* dalam pohon (Schonlau & Zou, 2020). Hal ini menghasilkan pohon-pohon yang tidak berkorelasi satu sama lain, sehingga ketika digabungkan, hasil prediksi menjadi lebih kuat dan andal. Selain itu, *Random Forest* juga memiliki mekanisme internal untuk mengestimasi error generalisasi melalui *out-of-bag* (OOB) error estimation, di mana sebagian data yang tidak digunakan untuk membangun pohon akan digunakan untuk menguji performa model (Setyadin et al., 2025).

Dapat disimpulkan *Random Forest* menjadi algoritma andalan dalam berbagai aplikasi machine learning karena kemampuannya menghasilkan model prediksi yang stabil, akurat, dan tahan terhadap *overfitting*. Dengan proses ensemble yang menggabungkan banyak pohon keputusan, *Random Forest* mampu memberikan solusi yang andal baik untuk masalah klasifikasi maupun regresi, dan sering dijadikan benchmark dalam pengembangan model prediktif modern.

### 3. METODE PENELITIAN

Metode dalam penelitian ini disusun untuk menggambarkan langkah-langkah, teknik, dan prosedur yang digunakan dalam pengumpulan serta analisis data. Proses tahapan dalam penelitian ini dapat dilihat pada *flowchart* yang terdapat pada Gambar 1.



**Gambar 1. Tahapan Penelitian**

#### Studi Literatur

Tahapan ini bertujuan untuk menelaah dan mengkaji teori-teori yang relevan dengan topik penelitian, dengan cara mengakses berbagai sumber seperti jurnal ilmiah, artikel, penelitian sebelumnya, serta referensi lainnya yang tersedia di internet.

#### Pengumpulan Data

Dataset yang dipakai adalah dataset penyakit jantung dari *UCI Machine Learning Repository* yang diambil dari *Kaggle* yang terdiri dari 303 data pasien.

#### Pra-pemrosesan Data

Tahapan pre-processing data digunakan untuk preparing dataset untuk kemudahan klasifikasi. Pengolahan data dilakukan dengan fitur *encoding* dan Normalisasi Data.

## 1. Normalisasi Data

Semua variabel numerik pada dataset, seperti usia, tekanan darah, kolesterol, dan lainnya, dinormalisasi menggunakan teknik *StandardScaler* untuk memastikan bahwa model tidak terpengaruh oleh perbedaan skala antar variabel. Proses normalisasi bertujuan untuk mengurangi bias yang dapat terjadi akibat skala yang berbeda antara variabel-variabel tersebut.

## 2. Encoding Data

Dataset ini juga memiliki beberapa variabel kategorikal seperti jenis kelamin (*sex*), tipe nyeri dada (*cp*), hasil EKG (*restecg*), dan variabel lainnya yang perlu diubah menjadi numerik agar bisa digunakan dalam model pembelajaran mesin.

Untuk melakukan encoding pada variabel kategorikal, digunakan teknik *One-Hot Encoding*. Teknik ini mengonversi setiap kategori dalam variabel menjadi kolom baru dengan nilai 0 atau 1, dimana 1 menandakan bahwa kategori tersebut ada, dan 0 menandakan tidak ada.

Sebagai contoh, variabel jenis kelamin yang memiliki dua kategori (Laki-laki dan Perempuan) akan diubah menjadi dua kolom baru, yaitu *sex\_1* (Laki-laki) dan *sex\_0* (Perempuan). Jika seorang pasien berjenis kelamin laki-laki, maka *sex\_1* akan bernilai 1, dan *sex\_0* bernilai 0.

*One-Hot Encoding* sangat berguna untuk menghindari masalah ordinal encoding, yang dapat menyebabkan model memandang hubungan antara kategori sebagai urutan atau peringkat (misalnya, menyangka bahwa '*Non-anginal Pain*' lebih rendah dari '*Typical Angina*').

## Pembagian Data

Dataset dikelompokkan menjadi dua bagian, yaitu data untuk testing (80%) dan data untuk training (20%). Pembagian ini menggunakan fungsi *train\_test\_split* dari library *scikit-learn*, yang memastikan bahwa model dilatih pada sebagian besar data dan diuji menggunakan data yang belum pernah dilihat sebelumnya.

## Pelatihan Model

Untuk memprediksi risiko penyakit jantung, model yang digunakan adalah *Random Forest Classifier*.

Model ini dilatih dengan menggunakan data pelatihan (*X\_train* dan *y\_train*). Proses pelatihan dilakukan menggunakan 100 estimator (*trees*) dengan pengaturan acak untuk meningkatkan keanekaragaman pohon keputusan yang dihasilkan, yang pada gilirannya meningkatkan kinerja model.

Setelah model dilatih, model dan skaler yang digunakan untuk normalisasi data disimpan menggunakan *pickle* dalam bentuk file `model_rf.pkl` untuk digunakan pada aplikasi web tanpa perlu pelatihan ulang setiap kali aplikasi dijalankan.

### **Pengujian dan Evaluasi Model**

Setelah model dilatih, evaluasi dilakukan untuk menguji kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya. Hasil prediksi model digunakan untuk memberikan penilaian apakah seseorang berisiko atau tidak terkena penyakit jantung.

Untuk memverifikasi kinerja model, evaluasi menggunakan matriks konfusi, akurasi, presisi, recall, dan F1-score dilakukan pada data pengujian.

### **Aplikasi Web untuk Prediksi**

Untuk mengakses model secara praktis, sebuah aplikasi web dibangun menggunakan *Flask*, sebuah framework Python untuk web development. Aplikasi web ini memungkinkan pengguna untuk memasukkan data pribadi seperti usia, jenis kelamin, tekanan darah, dan kolesterol, dan mendapatkan prediksi apakah mereka berisiko terkena penyakit jantung.

## **4. HASIL DAN PEMBAHASAN**

### **Pengumpulan Data**

Proses pengumpulan data dilakukan dengan memanfaatkan dataset penyakit jantung yang diambil dari *UCI Machine Learning Repository*. Dataset ini terdiri dari 303 data pasien yang mencakup berbagai variabel medis, seperti usia, jenis kelamin, tekanan darah, kolesterol, serta faktor medis lainnya yang berkaitan dengan risiko penyakit jantung.

Setiap data pasien telah diberi label target yang menunjukkan apakah pasien berisiko terkena penyakit jantung atau tidak. Dataset ini digunakan untuk pelatihan dan pengujian model prediksi penyakit jantung, dengan tujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi risiko penyakit jantung. Variabelnya terdiri dari:

1. *Age* – Usia (dalam tahun)
2. *Sex* – Jenis kelamin (1 = laki-laki; 0 = perempuan)
3. *Cp* – Tipe nyeri dada
4. *Trestbps* – Tekanan darah istirahat
5. *Chol* – Kolesterol serum (mg/dl)
6. *Fbs* – Gula darah puasa > 120 mg/dl (1 = ya; 0 = tidak)
7. *Restecg* – Hasil elektrokardiografi saat istirahat
8. *Thalach* – Detak jantung maksimum yang dicapai
9. *Exang* – Angina akibat olahraga

10. Oldpeak – Depresi ST yang diinduksi oleh olahraga
  11. Slope – Kemiringan segmen ST
  12. Ca – Jumlah pembuluh darah utama (0–3) yang diwarnai oleh fluoroskopi
  13. Thal – Hasil tes thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect)
- Target variabel (output) adalah target, yaitu label 1 untuk berisiko penyakit jantung dan 0 untuk tidak berisiko.

### Pra-pemrosesan Data

#### 1. Normalisasi Data

Semua variabel numerik di dataset, seperti usia, tekanan darah, dan kolesterol, dinormalisasi menggunakan *StandardScaler* agar semua variabel berada pada skala yang seragam.

#### 2. Encoding Data

Variabel kategorikal seperti jenis kelamin, tipe nyeri dada, dan hasil EKG diubah menjadi format numerik menggunakan *One-Hot Encoding*. Contoh variabel yang di *encoding* dapat dilihat pada tabel 1.

Tabel 1.

Jenis Kelamin	Encoding
Laki-laki	1
Perempuan	0

### Pembagian Data

Dataset terdiri dari 303 data pasien. Pembagian data dilakukan dengan menggunakan fungsi *train\_test\_split* dari *scikit-learn*, yang membagi dataset menjadi dua bagian, yaitu 80% data pelatihan dan 20% data pengujian.

Tabel 2.

Data Uji	Data Latih
61	242

### Pelatihan Model

Proses pelatihan model dimulai dengan mempersiapkan dataset yang mencakup fitur-fitur medis seperti usia, jenis kelamin, dan lainnya, serta variabel target yang menunjukkan apakah seseorang berisiko terkena penyakit jantung atau tidak. Data dibagi menjadi dua bagian: 80% untuk pelatihan dan 20% untuk pengujian. *Random Forest Classifier* adalah *algoritma ensemble learning* yang menggunakan banyak pohon keputusan (decision trees) untuk membuat prediksi. Setiap pohon keputusan dibangun menggunakan bootstrap sampling, di

mana subset data diambil secara acak dengan penggantian. Selanjutnya, setiap pohon keputusan dilatih pada subset data tersebut untuk membangun model yang dapat memprediksi apakah seorang pasien berisiko atau tidak terkena penyakit jantung. Pembentukan pohon keputusan didasarkan pada pembagian data pada setiap node, yang dipilih menggunakan *Gini Impurity* atau *Entropy*, yang mengukur ketidakpastian dalam pembagian data.

Setiap pohon keputusan dilatih secara mandiri, dan hasil prediksi dari semua pohon ini digabungkan melalui voting mayoritas untuk memberikan hasil akhir. Model ini dilatih dengan 100 *estimators* (pohon keputusan), di mana setiap pohon akan berkontribusi pada prediksi akhir. Gini Impurity, yang digunakan untuk memutuskan pembagian terbaik pada setiap node, dihitung dengan rumus berikut:

$$Gini(S) = 1 - \sum_{\{i=1\}}^{\{C\}} p_i^2$$

Proses ini memungkinkan model untuk belajar dari data training dan menghasilkan beberapa pohon yang dapat mengklasifikasikan data dengan baik.

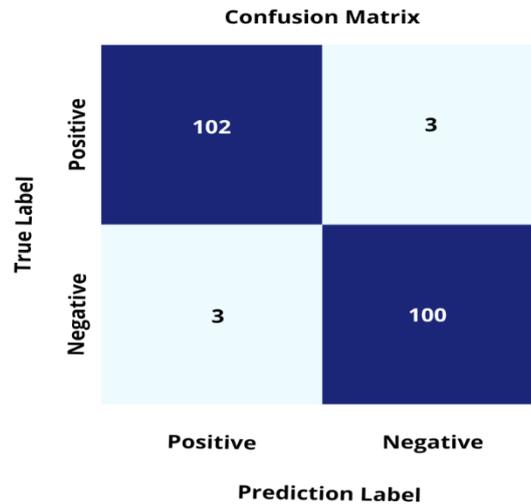
Setelah pohon keputusan selesai dibangun, model dapat digunakan untuk memprediksi risiko penyakit jantung pada data baru. Prediksi akhir dilakukan dengan mengambil hasil mayoritas suara dari semua pohon keputusan yang ada.

### **Pengujian dan Evaluasi Model**

Setelah model dilatih, tahap evaluasi dilakukan untuk mengukur performa model dalam memprediksi data yang asing. Salah satu cara yaitu menggunakan *Matriks Konfusi*, yang memberikan visualisasi yang jelas tentang prediksi model terhadap data uji. *Matriks Konfusi* digunakan untuk mengevaluasi kinerja model. Matriks memiliki empat komponen utama, yaitu :

1. *True Positive* (TP) : Kasus di mana model dengan benar memprediksi bahwa seseorang berisiko terkena penyakit jantung (positif).
2. *True Negative* (TN): Kasus di mana model dengan benar memprediksi bahwa seseorang tidak berisiko terkena penyakit jantung (negatif).
3. *False Positive* (FP): Kasus di mana model salah memprediksi bahwa seseorang berisiko terkena penyakit jantung padahal sebenarnya tidak berisiko (positif palsu).
4. *False Negative* (FN): Kasus di mana model salah memprediksi bahwa seseorang tidak berisiko terkena penyakit jantung padahal sebenarnya berisiko (negatif palsu).

Gambaran hasil dari confusion matrix dapat dilihat pada gambar 2.



**Gambar 2. Hasil Matriks Konfusi**

Selanjutnya akan dilakukan pengukuran performa model dengan menggunakan formula accuracy, precision, recall, dan f1-score dengan rumus dan hasil sebagai berikut.

1. *Accuracy* :

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

2. *Precision* :

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

3. *Recall* :

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

4. *F1-Score* :

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

Hasilnya dapat dilihat pada tabel 1.

**Tabel 3. Hasil Evaluasi**

Evaluasi	Hasil Evaluasi
<i>Accuracy</i>	0.985
<i>Recall</i>	0.971
<i>F1-Score</i>	0.985
<i>Precision</i>	1.000

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem deteksi dini penyakit jantung menggunakan algoritma *Random Forest* yang menggunakan dataset 303 data pasien dengan memanfaatkan teknik normalisasi data dan encoding variabel kategorikal untuk persiapan data dalam model prediksi. *Random Forest* terbukti efektif dalam mengklasifikasikan risiko penyakit jantung dengan tingkat akurasi yang sangat tinggi, yaitu 98,5%. Sistem ini dapat memberikan prediksi yang cepat dan akurat mengenai risiko penyakit jantung, yang akan sangat berguna untuk tenaga medis maupun individu dalam upaya deteksi dini penyakit jantung.

Sistem ini juga dirancang untuk dapat mengintegrasikan data secara *real-time*, yang memungkinkan pengguna untuk memperoleh informasi risiko secara langsung dan mendukung pengambilan keputusan medis. Dengan kemampuan ini, diharapkan dapat membantu memperkecil angka kematian akibat penyakit jantung yang dapat dicegah dengan deteksi lebih awal. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam bidang kesehatan, terutama dalam upaya meningkatkan sistem deteksi penyakit jantung dengan menggunakan teknologi kecerdasan buatan yang dapat diakses secara mudah dan efektif.

Dari hasil penelitian ini, disarankan untuk mengembangkan aplikasi lebih lanjut dengan menambahkan fitur-fitur tambahan yang dapat meningkatkan akurasi prediksi, seperti memasukkan data genetik atau kebiasaan hidup pasien. Selain itu, penerapan sistem ini pada populasi yang lebih besar dan lebih beragam sangat penting untuk menguji kemampuan generalisasi model agar lebih efektif. Integrasi sistem dengan perangkat medis lainnya juga perlu dilakukan untuk memberikan prediksi yang lebih komprehensif, yang akan mendukung pengambilan keputusan medis yang lebih baik. Selanjutnya, penelitian ini dapat diperluas dengan menggunakan dataset yang lebih bervariasi dari berbagai sumber untuk meningkatkan kualitas dan keberagaman data yang digunakan dalam model prediksi.

## DAFTAR REFERENSI

- Afandi, A. R., & Kurnia, H. (2023). Revolusi Teknologi: Masa Depan Kecerdasan Buatan (AI) dan Dampaknya Terhadap Masyarakat. *Academy of Social Science and Global Citizenship Journal*, 3(1), 9–13. <https://doi.org/10.47200/aossagcj.v3i1.1837>
- Alfajr, N. H., & Defiyanti, S. (2024). *METODE RANDOM FOREST DAN PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA)*. 12(3).
- All, K., Not, D., & Regression, L. (2006). *() Regression 1 / 13*. 50(x), 1–13.
- Atmojo, J. T., Ningrum, A. N., Handayani, R. T., Widiyanto, A., & Darmayanti, A. T. (2024). Artificial Intelligence dalam Praktik Kesehatan. *Jurnal Ilmiah Permas: Jurnal Ilmiah*

*STIKES Kendal*, 14(3), 1081–1088. <http://journal.stikeskendal.ac.id/index.php/PSKM>

- Elektronik, J., Udayana, I. K., Getrudis, S., Sadipun, I., Anom, G. N., Putra, C., Cs, M., Raya, J., & Unud, K. (2023). *Analisis Algoritma Random Forest Dalam Memprediksi Penyakit Jantung Koroner*. *Koroner*, 11(4), 2654–5101. <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- Haldy, J., & Meily Kurniawidjaja, L. (2024). Faktor Risiko Penyakit Kardiovaskular pada Pekerja : A Systematic Review. *Jurnal Kesehatan Masyarakat*, 8(1), 47–59.
- Jalan, K., Jamilah, N., Umar, S. M., & Ramba, J. (2024). *Penerapan Algoritma C4 . 5 dan Random Forest untuk Pemetaan*. 7, 326–339.
- Nurrohman, S. H., & Kurniawan, D. (2025). *Deteksi Dini Risiko Penyakit Jantung Koroner Menggunakan Algoritma Decision Tree dan Random Forest*. 6(4), 2582–2593. <https://doi.org/10.47065/bits.v6i4.7029>
- Nuryani, N., Purnama, B., Legowo, B., Utari, U., Riyatun, R., Suharno, S., Yuniyanto, M., & Lestari, W. (2023). Diseminasi Riset Kecerdasan Buatan Untuk Diagnosis Medis Berbasis Elektrokardiogram di Universitas Duta Bangsa. *SEMAR (Jurnal Ilmu Pengetahuan, Teknologi, Dan Seni Bagi Masyarakat)*, 12(2), 175. <https://doi.org/10.20961/semar.v12i2.75521>
- Pashar, I., Rahmi, Y. O., & Islamiyah, T. (2025). *MANUSIA DAN KESEHATAN Faktor-Faktor Risiko Terjadinya Penyakit Jantung Koroner Risk Factors For Coronary Heart Disease In Productive Age*. 8, 174–184.
- Rahmawati, A., Jayanti, K. D., Bisono, E. F., & Pangestuti, A. (2025). *Prediksi Penyakit Gagal Jantung Tahun 2025-2028 Rawat Inap di RSUD Gambiran Kota Kediri Institut Ilmu Kesehatan Bhakti Wiyata Kediri , Indonesia Data PKL mahasiswa IIK Bhakti Wiyata pada Triwulan 4 tahun 2023 di RSUD 10 besar penyakit rawat inap dengan 20*. 3.
- Sanhaji, G., & Hizbullah, A. I. (2023). Pemanfaatan Artificial Intelligence Dalam Bidang Kesehatan. *EDUSAINTEK: Jurnal Pendidikan, Sains Dan Teknologi*, 11(1), 234–242. <https://doi.org/10.47668/edusaintek.v11i1.999>
- Santosa, W. N., & Baharuddin, B. (2020). Penyakit Jantung Koroner dan Antioksidan. *KELUWIH: Jurnal Kesehatan Dan Kedokteran*, 1(2), 98–103. <https://doi.org/10.24123/kesdok.v1i2.2566>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- Setyadin, D., Winasis, R. H., & Triyono, G. (2025). *Deteksi Berita Palsu menggunakan Algoritma Random Forest Fake News Detection using the Random Forest Algorithm*. 14, 1142–1153.
- Sobron, M., & Lubis. (2021). Implementasi Artificial Intelligence Pada System Manufaktur Terpadu. *Seminar Nasional Teknik (SEMNASTEK) UISU*, 4(1), 1–7. <https://jurnal.uisu.ac.id/index.php/semnastek/article/view/4134>
- Trenggono, P. H., & Bachtiar, A. (2023). Peran Artificial Intelligence Dalam Pelayanan

Kesehatan : a Systematic Review. *Jurnal Ners*, 7(1), 444–451.  
<https://doi.org/10.31004/jn.v7i1.13612>

Wahyudi, T. (2023). Studi Kasus Pengembangan dan Penggunaan Artificial Intelligence (AI) Sebagai Penunjang Kegiatan Masyarakat Indonesia. *Indonesian Journal on Software Engineering (IJSE)*, 9(1), 28–32. <https://ejournal.bsi.ac.id/ejurnal/index.php/ijse>

Wijaya, M., Tumbelaka, L. G., & Bandana, V. (2025). *KORONER DAN ARITMIA MELALUI MEDIA SOSIAL INSTAGRAM DAN YOUTUBE*. 03(01), 1–10.