

PENGEMBANGAN APLIKASI PREDIKSI RISIKO PENYAKIT JANTUNG BERBASIS MODEL KLASIFIKASI RANDOM FOREST MENGGUNAKAN FRAMEWORK STREAMLIT

Fitra Salam S. Nagalay^{1*}, Desi Rahma Aryanti², Meri Liandani³

Fakultas Sains dan Teknologi, Universitas Wira Buana^{1,2}

Fakultas Ilmu Kesehatan, Universitas Wira Buana³

fitrasalamsn@gmail.com¹, desirahmaaryanti@gmail.com², meriliandani18@gmail.com³

Email Korespondensi: fitrasalamsn@gmail.com *

ABSTRACT

Coronary Heart Disease (CHD) is one of the leading causes of mortality worldwide, making early risk detection crucial. The advancement of machine learning technology offers significant potential in developing accurate and accessible prediction tools. This study aims to develop a web-based calculator application to predict the risk of CHD. The method used is a Random Forest classification model trained on the UCI Heart Disease dataset, which was chosen after being compared with several other algorithms. The model was then implemented into an interactive web application using the Streamlit framework. The research process included data preprocessing, model training, performance evaluation, and user interface development. The evaluation results showed that the Random Forest model achieved the best performance with an accuracy of 88.52%. The resulting application successfully provides a real-time risk percentage visualization via an intuitive gauge chart, making it easily understandable for lay users. This application is expected to serve as an effective preliminary screening tool to increase public awareness of CHD risk.

Keywords: *Machine Learning, Random Forest, Coronary Heart Disease (CHD), Web Application, Streamlit.*

ABSTRAK

Penyakit Jantung Koroner (PJK) merupakan salah satu penyebab utama kematian di seluruh dunia, sehingga deteksi dini faktor risiko menjadi sangat krusial. Perkembangan teknologi *machine learning* menawarkan potensi besar dalam pengembangan alat bantu prediksi yang akurat dan mudah diakses. Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi kalkulator berbasis web untuk memprediksi risiko PJK. Metode yang digunakan adalah model klasifikasi *Random Forest* yang dilatih menggunakan *dataset* Penyakit Jantung dari UCI setelah dibandingkan dengan beberapa algoritma lainnya. Model tersebut kemudian diimplementasikan ke dalam aplikasi web interaktif menggunakan *framework* Streamlit. Hasil evaluasi menunjukkan model *Random Forest* mencapai performa terbaik dengan tingkat akurasi sebesar 88.52%. Aplikasi yang dihasilkan mampu memberikan visualisasi persentase risiko secara *real-time* melalui *gauge chart* yang intuitif sehingga mudah dipahami oleh pengguna awam. Aplikasi ini diharapkan dapat menjadi alat bantu skrining awal yang efektif untuk meningkatkan kesadaran masyarakat akan risiko PJK.

Kata Kunci: *Machine Learning, Random Forest, Penyakit Jantung Koroner (PJK), Aplikasi Web, Streamlit*

PENDAHULUAN

Penyakit kardiovaskular, khususnya yang berkaitan dengan jantung, telah menjadi tantangan kesehatan global dan penyebab utama mortalitas di berbagai negara, termasuk Indonesia (Fauzan et al., 2025). Deteksi dini dan pengelolaan faktor-faktor risiko merupakan kunci utama dalam upaya pencegahan. Faktor risiko ini mencakup usia, jenis kelamin, tekanan darah, kadar kolesterol, dan status merokok (Rosjidi et al., 2021).

Secara tradisional, penilaian risiko dilakukan oleh tenaga medis profesional. Namun, kemajuan teknologi, khususnya *machine learning*, telah membuka peluang untuk melakukan analisis prediktif di bidang kesehatan dengan akurasi yang tinggi. Algoritma *machine learning* mampu mempelajari pola kompleks dari data historis pasien untuk membuat prediksi (Thaariq et al., 2024).

Beberapa penelitian sebelumnya telah membuktikan efektivitas algoritma *machine learning*. Penelitian oleh Inayah dan Ramfli menunjukkan bahwa algoritma *Random Forest* memiliki kinerja yang unggul karena kemampuannya menangani data dalam jumlah besar dan mengatasi *overfitting* (Inayah & Ramli, 2024). Meskipun model prediktif telah banyak dikembangkan, implementasinya dalam bentuk aplikasi yang ramah pengguna

(*user-friendly*) masih menjadi tantangan (Mahendra et al., 2025). Oleh karena itu, penelitian ini berfokus pada pengembangan sebuah aplikasi web yang tidak hanya akurat tetapi juga interaktif. Aplikasi ini dibangun menggunakan model klasifikasi *Random Forest* yang terbukti paling akurat dari beberapa model yang diuji, dan diimplementasikan dengan *framework* Streamlit. Tujuan utama dari penelitian ini adalah menyediakan alat bantu skrining awal yang dapat membantu individu memahami tingkat risiko penyakit jantung mereka.

METODE

Metodologi penelitian ini disusun secara sistematis, mulai dari, pengumpulan data, pra-pemrosesan data, pemodelan, hingga pengembangan aplikasi web.

a. Dataset

Data yang digunakan adalah *dataset* "Heart Disease UCI" dari *UCI Machine Learning Repository*, yang terdiri dari 303 sampel dan 14 atribut. Atribut target dalam dataset ini adalah variabel 'target', yang mengindikasikan ada atau tidaknya penyakit jantung pada pasien (1 = ada, 0 = tidak ada) (Amelia, 2023).

b. Pra-pemrosesan Data

Pada tahap ini data akan dilakukan proses preprocessing yang melibatkan

pengelolaan data awal, seperti pembersihan data (*cleaning data*). Langkah ini bertujuan untuk Menilai bahwasannya data yang dipakai pada pelatihan memiliki kualitas yang bagus tidak ditemukan nilai yang hilang (*missing values*) pada *dataset* ini, sehingga tahap penanganan data hilang tidak diperlukan (Sudrajat & Cholid, 2023). Data kemudian dibagi menjadi data latihan (*training set*) dan data uji (*testing set*) dengan proporsi 80:20 untuk mempersiapkan proses pemodelan (Indayana et al., 2025).

c. Pemodelan Machine Learning

Pada tahap pemodelan, dilakukan evaluasi komparatif terhadap beberapa algoritma klasifikasi untuk menemukan model dengan performa terbaik. Proses perbandingan ini merupakan langkah krusial dalam penelitian *machine learning*, karena tidak ada satu algoritma tunggal yang superior untuk semua jenis dataset; pemilihan model yang optimal harus didasarkan pada kinerja empiris pada data yang spesifik (Wijiyanto et al., 2024). Untuk tujuan ini, penelitian ini mengevaluasi serangkaian algoritma klasifikasi yang umum digunakan dan telah terbukti efektif dalam berbagai studi, yang mencakup model statistik probabilitik seperti *Logistic Regression* dan *Naive Bayes*, algoritma berbasis

instans seperti *K-Nearest Neighbors (KNN)*, dan model yang bekerja dengan mencari batas keputusan optimal seperti *Support Vector Machine (SVM)*. Selain itu, diuji pula model berbasis pohon keputusan (*Decision Tree Classifier*) beserta metode *ensemble*-nya yaitu *Random Forest Classifier*, yang dikenal mampu meningkatkan stabilitas dan akurasi dengan menggabungkan hasil dari banyak pohon keputusan. Kumpulan algoritma ini dipilih karena secara kolektif merepresentasikan pendekatan yang beragam dalam tugas klasifikasi dan kinerjanya sering dianalisis secara bersamaan dalam studi perbandingan (Das et al., 2023). Setiap model dilatih menggunakan data latihan dan dievaluasi akurasi pada data uji untuk menentukan model dengan kinerja prediktif terbaik secara objektif.

d. Pengembangan Aplikasi

Model dengan *algoritma* terbaik disimpan ke dalam file yang berekstensi *.pkl* menggunakan *library* *Pickle*. Kemudian Aplikasi web dikembangkan menggunakan *framework* *Streamlit* di *Python*. *Streamlit* memungkinkan pembuatan antarmuka pengguna yang interaktif dengan cepat, di mana pengguna dapat memasukkan data melalui *widget* seperti *text input*, *slider*, dan *selectbox*.

HASIL

Pada penelitian ini dataset yang sudah di lakukan pra-pemrosesan data dan juga di lakukan pembersihan data. Maka data tersebut akan dilakukan pemodelan dengan beberapa algoritma untuk dibandingkan berikut menunjukkan hasil perbandingan akurasi dari setiap model yang diuji pada dataset.

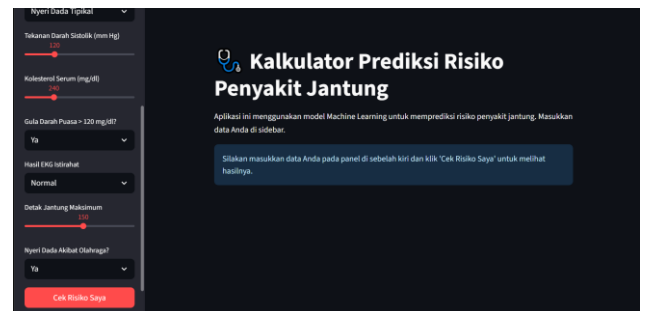
Tabel 1. Perbandingan Akurasi Algoritma

Model	Akurasi (%)
Logistic Regression	86.89
Support Vector Machine	86.89
K-Nearest Neighbors	68.85
Decision Tree	80.33
Random Forest	88.52
Naive Bayes	86.89

Hasil perbandingan dari tiap model yang telah di uji dapat dilihat Dari hasil pada Tabel 1, terlihat bahwa model Random Forest menunjukkan performa tertinggi dengan akurasi mencapai 88.52%. Model lain seperti Logistic Regression, SVM, dan Naive Bayes juga menunjukkan kinerja yang baik dengan akurasi identik sebesar 86.89%. Namun, keunggulan Random Forest tidak hanya terletak pada akurasi yang sedikit lebih tinggi, tetapi juga pada robustitasnya sebagai metode *ensemble* yang mampu mengurangi

overfitting dan memberikan wawasan mengenai fitur yang paling berpengaruh dalam pengambilan keputusan. Oleh karena itu, model Random Forest dipilih sebagai basis untuk pengembangan aplikasi kalkulator risiko ini.

Setelah menyimpan model dengan algoritma terbaik maka selanjutnya pengembangan aplikasi. Aplikasi yang dihasilkan memiliki antarmuka yang mudah digunakan. Pengguna memasukkan nama dan 9 parameter kesehatan pada sidebar. Tampilan aplikasi kalkulator prediksi dapat dilihat pada Gambar 1.



Gambar 1. Tampilan Aplikasi Kalkulator Prediksi Penyakit Jantung

Setelah pengguna memasukkan data kemudian menekan tombol "Cek Risiko Saya", aplikasi akan menampilkan hasil analisis yang dipersonalisasi dengan nama pengguna, diikuti dengan visualisasi *gauge chart* yang menunjukkan persentase risiko (Jafari et al., 2023).



Gambar 2. Visualisasi Hasil Prediksi dengan Gauge Chart

Pendekatan ini memberikan pemahaman yang lebih bernuansa kepada pengguna mengenai sejauh mana tingkat risiko mereka. Meskipun pengujian menunjukkan bahwa model ini memiliki batas prediksi probabilitas maksimal sekitar 70%, angka ini sudah secara tegas masuk ke dalam kategori "Risiko Tinggi" pada visualisasi terlihat jelas bahwasannya aplikasi memberikan peringatan yang jelas dan memadai bagi pengguna.

PEMBAHASAN

Kinerja unggul model Random Forest dengan akurasi 88.52% sejalan dengan temuan pada penelitian-penelitian sebelumnya. Sebagai metode *ensemble*, Random Forest menggabungkan prediksi dari banyak *decision tree* untuk menghasilkan keputusan akhir, yang secara efektif mengurangi varians dan risiko *overfitting*. Keunggulannya dibandingkan model tunggal seperti Decision Tree (80.33%) sangat terlihat dalam hasil penelitian ini. Performa

Random Forest yang sedikit lebih baik dari Logistic Regression dan SVM (86.89%) juga menunjukkan kemampuannya dalam menangkap hubungan non-linear yang kompleks antar fitur dalam data medis.

Pengembangan aplikasi ini menjawab tantangan utama dalam penelitian *machine learning* di bidang kesehatan, yaitu bagaimana mentransformasikan model prediktif yang kompleks menjadi alat yang praktis dan dapat diakses. Dengan menggunakan Streamlit, proses implementasi menjadi lebih efisien. Penggunaan visualisasi seperti *gauge chart* untuk menampilkan persentase risiko adalah keputusan desain yang krusial. Dibandingkan hanya memberikan output biner "berisiko" atau "tidak berisiko", penyajian dalam bentuk persentase memberikan spektrum pemahaman yang lebih luas bagi pengguna dan dapat memotivasi tindakan preventif bahkan pada tingkat risiko sedang. Ini sejalan dengan tujuan penelitian untuk menciptakan alat bantu skrining yang edukatif.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan, dapat disimpulkan bahwa penelitian ini telah berhasil mengembangkan sebuah aplikasi

web fungsional untuk prediksi risiko penyakit jantung, yang secara efektif menjawab tujuan utama untuk menciptakan alat bantu skrining berbasis teknologi. Melalui perbandingan sistematis terhadap enam algoritma klasifikasi, model Random Forest terbukti menjadi metode yang paling unggul dengan mencapai akurasi 88.52%, sehingga menjadi dasar dari mesin prediksi aplikasi ini. Aplikasi yang dihasilkan mampu memberikan umpan balik yang dipersonalisasi dan mudah dipahami kepada pengguna melalui antarmuka interaktif dan visualisasi data berupa *gauge chart*, yang berhasil memenuhi tujuan untuk menciptakan sebuah alat yang *user-friendly* dan informatif untuk deteksi dini risiko penyakit jantung.

DAFTAR PUSTAKA

- Fauzan R, Zahrina, Siregar SR. Peningkatan Kesadaran Masyarakat terhadap Risiko Gagal Jantung melalui Kegiatan Penyuluhan. *Auxilium J Pengabdian Kesehatan*. 2025;3(2):7–11.
- Rosjidi CH, Isro'in L, Wahyuni NS. *Buku Ajar Asuhan Keperawatan pada Pasien dengan Gangguan Sistem Kardiovaskular*. Ponorogo: Myria Publisher; 2021.
- Thaariq MA, Baskara MD, Chaniago RA, Christin D, Ernawati I. *Systematic Literature Review: Analisis Penerapan Kecerdasan Buatan Dalam Bidang Kesehatan*. InProsiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya 2024 Aug 19 (Vol. 5, No. 1, pp. 168-173).
- Inayah K, Ramli K. Analisis Kinerja Intrusion Detection System (IDS) Suricata Menggunakan Metode Klasifikasi Random Forest. *J Teknol Inf dan Ilmu Komput*. 2024;11(4):867–76.
- Mahendra GS, Jailani AK, Agus M, Kelvin K, Sari PA, Wardani DK, Judijanto L. *Smart Computing IoT & Machine Learning*. PT. Sonpedia Publishing Indonesia; 2025 Aug 4.
- Indayana NS, Arman D, Okmayura F, Anjani SP, Dayani FN, Farhan M, Faturrahman A. PEMODELAN MACHINE LEARNING DENGAN ALGORITMA RANDOM FOREST DALAM MEMREDIKSI RISIKO STROKE. *Journal of Software Engineering and Information System (SEIS)*. 2025 Aug 20.
- Amelia Y. Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung. *IDEALIS Indones J Inf Syst*. 2023;6(2):220–5.
- Sudrajat W, Cholid I. K-NEAREST NEIGHBOR (K-NN) UNTUK PENANGANAN MISSING VALUE PADA DATA UMKM. *J Rekayasa Sist Inf dan Teknol*. 2023;1(2):54–63.
- Wijiyanto, Pradana AI, Sopingi S. Perbandingan Data Untuk Memprediksi Ketepatan Studi Berdasarkan Atribut Keluarga Menggunakan Machine Learning. *JIKA (Jurnal Inform*. 2024;8(2):221–8.
- Das S, Bhattacharyya K, Sarkar S. Performance Analysis of Logistic Regression, Naive Bayes, KNN, Decision Tree, Random Forest and SVM on Hate Speech Detection

from Twitter. *Int Res J Innov Eng Technol.* 2023;7(3):24–8.

Jafari M, Maleki MR, Salmasnia A. A high-dimensional control chart for monitoring process variability under gauge imprecision effect. *Prod Eng.* 2023;17:547–64.