

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Jantung merupakan organ terpenting pada tubuh manusia karena mempunyai fungsi memompa darah ke seluruh tubuh dalam memenuhi kebutuhan oksigen serta jaringan nutrisi. Apabila fungsi jantung terganggu, misalnya kemampuan memompanya menurun, hal ini dapat menimbulkan gangguan serius seperti serangan jantung. Penyakit jantung dapat dikatakan sebagai penyakit yang jumlah kasusnya cukup tinggi dan menjadi penyebab utama angka kematian, terutama di Indonesia. Kondisi ini terjadi ketika jantung tidak mampu bekerja secara optimal dalam memompa darah ke seluruh tubuh (Ryfai dkk., 2022).

Berdasarkan laporan *World Health Organization* (WHO) tahun 2021, tercatat sekitar 17,8 juta kematian terjadi setiap tahun, dengan lebih dari 80% di antaranya disebabkan oleh penyakit jantung dan stroke. Di Indonesia sendiri, penyakit jantung menempati posisi sebagai penyebab kematian tertinggi kedua setelah stroke. Menurut data Pusat Analisis Keparlemenan Badan Keahlian Setjen DPR RI, jumlah kasus penyakit jantung pada tahun 2021 tercatat sebanyak 12,93 juta kasus. Lalu pada tahun 2022 kasusnya mengalami peningkatan hingga mencapai 15,5 juta (Lestari, 2023). Peningkatan jumlah penderita ini menggambarkan bahwa penyakit jantung masih menjadi tantangan besar dalam bidang kesehatan masyarakat.

Seiring kemajuan teknologi, analisis data dapat dimanfaatkan untuk mendukung deteksi dini dan klasifikasi risiko penyakit jantung. Dengan *data mining*, informasi penting dari kumpulan data medis dapat diekstraksi untuk memahami pola dan hubungan antarvariabel (Damuri dkk., 2021). Salah satu pendekatan yang umum digunakan adalah klasifikasi, yaitu dapat membantu menentukan apakah seorang pasien termasuk dalam kelompok berisiko atau tidak berisiko terkena penyakit jantung (Rashad dkk., 2023). Pendekatan ini mendukung tenaga medis dalam merencanakan tindakan pencegahan yang tepat.

Dalam penelitian ini dipilih algoritma *Random Forest*, yaitu metode *ensemble learning* berbasis pohon keputusan (*decision tree*) yang memadukan hasil prediksi dari sejumlah pohon sehingga mampu menghasilkan prediksi yang lebih stabil dan akurat (Amalia dkk., 2022). *Random Forest* juga memiliki keunggulan dalam mengolah data dengan jumlah variable yang besar, serta cukup tangguh terhadap *outlier* dan *missing*. Selain itu, algoritma *Random Forest* juga dapat melihat *Feature Importance* (fitur paling berpengaruh) sehingga dapat meningkatkan efisiensi model (Ernawati dkk., 2025).

Untuk mendapatkan kualitas yang baik dalam penelitian, peneliti juga menerapkan metode *Monte Carlo Cross Validation* (MCCV). Menurut (Kuncoro, 2021), metode ini merupakan teknik validasi yang membagi data latih dan data uji secara acak dalam beberapa kali pengulangan. Setiap hasil evaluasi dari pengulangan digabungkan secara rata-rata, sehingga estimasi kinerja model menjadi lebih stabil dan risiko bias dapat diminimalkan. Bias yang dimaksud adalah pembagian data yang menghasilkan komposisi data latih dan data uji tidak seimbang.

Beberapa penelitian sebelumnya telah dilakukan untuk memprediksi penyakit jantung menggunakan berbagai metode *machine learning*. Penelitian yang dilakukan oleh (Ayu & Maharani, 2023) dengan judul “Penerapan Metode *Naïve Bayes* Dalam Memprediksi Penyakit Jantung” menggunakan algoritma *Naïve Bayes* untuk melakukan proses klasifikasi penyakit jantung dan menghasilkan akurasi sebesar 81%. Penelitian lain yang dilakukan oleh (Muzakki dkk., 2024) dengan judul “Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan *RapidMiner*” menunjukkan bahwa algoritma C4.5 mampu menghasilkan akurasi sebesar 65,25%, *recall* sebesar 70,87%, dan *presisi* sebesar 62,46%. Selain itu, penelitian yang dilakukan oleh (Hidayat dkk., 2023) dengan judul “Klasifikasi Penyakit Jantung Menggunakan *Random Forest Classifier*” menunjukkan bahwa metode *Random Forest* dapat menghasilkan performa yang baik dengan akurasi sebesar 94%.

Berdasarkan uraian tersebut, diperlukan pengembangan model prediksi penyakit jantung yang mampu memberikan performa optimal dilihat dari aspek

akurasi, *presisi*, *recall*, maupun *f1-score*. Oleh karena itu, penulis berminat melakukan penelitian dengan judul “**PENERAPAN METODE MONTE CARLO CROSS VALIDATION PADA ALGORITMA RANDOM FOREST DALAM KLASIFIKASI RISIKO SERANGAN JANTUNG**”.

## **1.2 Rumusan Masalah**

Merujuk pada latar belakang, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana rata-rata hasil pengukuran performa algoritma *Random Forest* berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-score* setelah dilakukan pembagian data menggunakan metode *Monte Carlo Cross Validation*?
2. Fitur medis apa yang paling berpengaruh dalam mengklasifikasi serangan jantung?

## **1.3 Tujuan**

Adapun tujuan dari rumusan masalah diatas:

1. Untuk mengetahui rata-rata performa algoritma *Random Forest* berdasarkan metrik akurasi, *presisi*, *recall*, dan *F1-score* setelah dilakukan pembagian data menggunakan metode *Monte Carlo Cross Validation*
2. Mengidentifikasi fitur medis yang paling berpengaruh dalam mengklasifikasi serangan jantung

## **1.4 Batasan Masalah**

Agar penelitian lebih terarah dan fokus, maka batasan masalah ditetapkan sebagai berikut:

1. *Dataset* yang digunakan dalam penelitian ini adalah Kumpulan Data Serangan jantung yang bersumber dari *Kaggle* sebanyak 1319 data, yang dapat diakses melalui: <https://www.kaggle.com/datasets/fatemeahmammadinia/heart-attack-dataset-tarik-a-rashid?resource=download>
2. Terdiri dari beberapa fitur medis, antara lain *Age* (usia), *Gender* (jenis kelamin), *Heart Rate* (detak jantung), *Systolic Blood Pressure* (tekanan darah

atas), *Diastolic Blood Pressure* (tekanan darah bawah), *Blood Sugar* (kadar gula), *CK-MB* (enzim keratin), dan *Troponin* dan *Result* dengan target 0 = tidak mengalami serangan jantung, 1 = mengalami serangan jantung.

3. Algoritma klasifikasi yang digunakan pada penelitian ini yaitu algoritma *Random Forest*, tanpa membandingkan dengan algoritma klasifikasi lainnya.
4. Model diuji dengan metode *Monte Carlo Cross Validation* dengan dilakukannya proses pengulangan sampai iterasi stabil.
5. Menggunakan skenario uji 80:20, 70:30, dan 60:40 dengan 1000 iterasi di setiap skenarionya.
6. Menghitung *Feature Importance* (fitur paling berpengaruh) pada model.
7. Evaluasi performa model dibatasi pada parameter akurasi, *presisi*, *recall*, dan *F1-score*.
8. *Tools* yang dipakai pada penelitian ini adalah *Google Colaboratory* (*Google Colab*) dengan bahasa pemrograman *Python*.

### **1.5 Manfaat**

Dari hasil penelitian ini, diharapkan dapat diperoleh manfaat sebagai berikut:

1. Memberikan pemahaman lebih mendalam terkait penggunaan teknik *Monte Carlo Cross Validation* dalam algoritma *Random Forest*.
2. Dapat dijadikan referensi dan acuan dalam penelitian lanjutan.
3. Dapat menjadi dasar untuk membangun sistem deteksi dini risiko penyakit jantung yang lebih akurat dan dapat diimplementasikan dalam praktik nyata.