

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Penyakit jantung merupakan penyebab utama kematian di dunia. Menurut WHO (World Health Organization) dan data Riskesdas (Riset Kesehatan Dasar), angka kematian terus meningkat akibat gaya hidup tidak sehat seperti pola makan yang buruk, kurang olahraga, merokok, serta stress tinggi. Pada tahun 2022, Indonesia mencatat 15,5 juta kasus penyakit jantung yang menghabiskan 21% total biaya BPJS Kesehatan. Penyakit ini menyerang semua kelompok umur dan menjadi beban kesehatan nasional yang sangat besar[1].

Pengembangan sistem untuk mendeteksi penyakit jantung pada usia dini sangat penting, terutama mengingat banyaknya penderita yang berada pada rentang usia produktif. Sistem ini memungkinkan intervensi medis yang lebih cepat dan tepat, serta mengurangi biaya yang digunakan terkait dengan penyakit jantung. Pemanfaatan algoritma mesin learning adalah salah satu pendekatan teknologi yang dapat membantu deteksi dini penyakit jantung. Metode ini memungkinkan analisis data dilakukan secara otomatis, yang dapat meningkatkan ketepatan proses diagnosis medis[2].

Machine learning adalah salah satu teknologi saat ini yang dapat digunakan untuk mendeteksi penyakit. Machine learning sendiri bukanlah teknologi yang dapat menganalisis data dengan benar sepenuhnya dan kemudian membuat kesimpulan[3].

Pada penelitian Analisis perbandingan algoritma klasifikasi untuk penyakit jantung yang dilakukan oleh Jalis Dwi Muthohhar dan Agus Prihanto memberikan hasil bahwa, Algoritma *Random forest* merupakan model terbaik dalam menganalisis penyakit jantung dengan menggunakan *hyperparameter tuning grid search*[3]. Pada penelitian Perbandingan Metode Machine Learning untuk Mendeteksi Penyakit Jantung yang dilakukan oleh Yutri Amelia menyatakan bahwa algoritma XGBoost memperoleh nilai akurasi, sensitifitas, presisi dan *f1-score* tertinggi untuk menganalisis penyakit jantung, masing-masing

nilainya 95,08%, 93,80%, 96,80% dan 95,20%. sebesar 91.13% dan positif sebesar 93.85%[4]. Kemudian pada penelitian Studi Komparatif Model Klasifikasi Kerentanan Penyakit Jantung Menggunakan Algoritma Machine Learning yang dilakukan oleh Wiji Lestari, dan Sri Sumarlinda menyatakan bahwa algoritma *Logistic Regression* mempunyai nilai tertinggi 81,00%[5].

Pada penelitian ini peneliti akan membandingkan algoritma Random Forest, Logistic Regression, dan XGBoost untuk menemukan algoritma yang paling tepat dalam menganalisis data penyakit jantung.

### **1.2 Rumusan Masalah**

1. Berapa akurasi model algoritma Random Forest, Logistic Regression, dan XGBoost dalam menganalisis penyakit jantung ?
2. Bagaimana perbedaan model sebelum dilakukan hyper parameter dan setelah dilakukan hyper parameter menggunakan metrik ROC-AUC sebagai metrik utama?

### **1.3 Batasan Masalah**

1. Penelitian ini menggunakan dataset penyakit jantung yang bersumber dari platform Kaggle, sehingga hasil penelitian bersifat umum dan tidak merepresentasikan kondisi pasien secara spesifik di wilayah tertentu.
2. Variabel input (fitur) yang digunakan meliputi: age, sex, chest pain type, resting bps, fasting blood sugar, resting eeg, maximum heart rate, exercise angina, oldpeak, st slope, dan kolom yang tidak digunakan (dihapus) adalah unnamed: 0, dan cholesterol, sedangkan variabel output (target) adalah target, yang bernilai 0 untuk tidak berpotensi penyakit jantung dan 1 untuk berpotensi penyakit jantung.
3. Model yang di uji pada penelitian ini yaitu Random Forest, Logistic Regression dan XGBoost

### **1.4 Tujuan Penelitian**

1. Membandingkan performa tiga algoritma machine learning, yaitu Random Forest, Logistic Regression, dan XGBoost Classifier dalam menganalisis

- penyakit jantung berdasarkan dataset publik yang diperoleh dari Kaggle.
2. Mengidentifikasi algoritma dengan tingkat akurasi dan performa terbaik berdasarkan metrik evaluasi seperti accuracy, precision, recall, F1-score, ROC-AUC dan kurva ROC.
  3. Menentukan faktor-faktor fitur yang paling berpengaruh terhadap hasil analisis penyakit jantung pada setiap model yang diuji.
  4. Memberikan rekomendasi algoritma paling efektif yang dapat digunakan pada kasus analisis serupa di bidang kesehatan, khususnya untuk menganalisis penyakit jantung berbasis data publik.

### **1.5 Manfaat Penelitian**

1. Memberikan kontribusi teoritis terhadap pengembangan ilmu pengetahuan di bidang machine learning dan data science, khususnya dalam penerapan algoritma untuk analisis data kesehatan.
2. Menjadi referensi akademik bagi peneliti selanjutnya dalam melakukan perbandingan performa berbagai algoritma machine learning pada kasus analisis berbasis data publik.

### **1.6 Sistematika Penulisan**

Sistematika penulisan yang digunakan untuk menyusun skripsi ini terdapat 5 bab. Sistematika penulisan skripsi sebagai berikut :

BAB I PENDAHULUAN, pada bab ini membahas latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan yang digunakan dalam menyusun skripsi ini.

BAB II TINJAUAN PUSTAKA, pada bab ini menjelaskan studi literatur dan dasar teori yang di gunakan dalam mendukung penelitian ini

BAB III METODE PENELITIAN, pada bab ini berisi objek penelitian, alur penelitian serta alat dan bahan yang digunakan.

BAB IV HASIL DAN PEMBAHASAN, bab ini menjelaskan pembahasan dan hasil yang diperoleh setelah melakukan penelitian.

BAB V PENUTUP, bab ini berisi saran dan kesimpulan dari hasil penelitian yang telah dilakukan.

