

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung adalah salah satu masalah kesehatan utama yang menyebabkan kematian tertinggi di dunia. Faktor-faktor risiko seperti tekanan darah tinggi, diabetes, obesitas, gaya hidup tidak sehat, merokok, dan faktor genetic mendeteksi dan memprediksi penyakit jantung. Menurut data dari *World Health Organization WHO*, lebih dari 17 juta orang meninggal setiap tahun akibat penyakit ini [1]. Hal ini menjadikan penyakit jantung sebagai beban berat bagi kesehatan global dan kondisi yang dimana sering kali disebabkan oleh gaya hidup tidak sehat, seperti pola makan yang buruk, kurang aktivitas fisik, merokok, dan tekanan darah tinggi. Gejala penyakit jantung sering kali tidak tampak pada tahap awal, sehingga banyak kasus baru terdeteksi setelah mencapai kondisi yang parah dan mendorong fokus penelitian untuk mengembangkan strategi pencegahan yang efektif [2].

Dengan kemajuan teknologi dalam bidang data mining dan *machine learning*, berbagai pendekatan telah dimanfaatkan untuk pemanfaatan *machine learning* dalam bidang kesehatan semakin populer, khususnya dalam hal prediksi dan deteksi dini penyakit, termasuk penyakit jantung dan memperoleh tingkat prediksi yang tinggi dalam mendiagnosis penyakit jantung dan pembelajaran mesin, bagian dari kecerdasan buatan, memungkinkan komputer untuk belajar dan membuat keputusan tanpa harus deprogram ulang secara eksplisit. Salah satu bagian penting dari pengobatan atau tindakan medis adalah pengambilan keputusan dan proses prediksi pada suatu hal yang menjadi fokus seperti pendeteksian penyakit, namun prediksi medis biasanya merupakan proses yang sangat kompleks dan sulit dilakukan jika tak dapat mengetahui metode yang tepat dan terbaik dalam memberikan solusinya. Tantangan yang dihadapi oleh organisasi kesehatan adalah memprediksi pasien dengan benar, prediksi yang buruk dapat menyebabkan konsekuensi yang mendatangkan malapetaka yang

kemudian tidak dapat diterima. Dua algoritma yang digunakan dalam penelitian ini adalah *Random Forest* dan *Decision Tree*. Dua metode yang populer dalam klasifikasi dan prediksi dalam domain kesehatan. *Decision Tree* dikenal karena interpretasi nya yang mudah melakukan proses klasifikasi atau prediksi yang sederhana, sedangkan *Random Forest* adalah pengembangan dari *Decision Tree* yang menggabungkan beberapa keputusan pohon untuk meningkatkan akurasi dan mengurangi risiko *overfitting* [1].

Manfaat yang diharapkan dari penelitian ini adalah bersifat multidimensi, mencakup aspek praktis maupun teoritis. Pada tingkat praktis, penelitian ini bertujuan menghasilkan alat bantu prediksi yang dapat diintegrasikan ke dalam praktik klinis sehari-hari, membantu tenaga medis dalam proses *screening* dan diagnosis awal terkena penyakit jantung. Pada tingkat teoretis, penelitian ini berkontribusi pada pemahaman yang lebih dalam tentang hierarki kepentingan berbagai parameter kesehatan dalam konteks prediksi penyakit jantung. Dengan memanfaatkan analisis perbandingan dua algoritma yaitu *Random Forest* dan *Decision Tree* kekayaan informasi dalam dataset yang digunakan, penelitian ini diposisikan untuk menghasilkan model prediksi yang tidak hanya akurat secara statistik tetapi juga aplikatif dalam konteks klinis.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk membandingkan performa algoritma *Random Forest* dan *Decision Tree* dalam memprediksi resiko terkena penyakit jantung. Perbandingan ini dilakukan berdasarkan beberapa metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*, agar dapat mengetahui algoritma mana yang lebih efektif dan efisien dalam konteks data medis yang digunakan.

1.2 Ruang Lingkup

- a. Data yang diprediksi merupakan data riwayat terkena penyakit jantung, data yang digunakan berasal dari repositori web *kaggle*.
- b. Menganalisis dan membandingkan tingkat akurasi dari kedua algoritma yaitu *random forest* dan *decision tree* dalam memprediksi resiko terkena penyakit jantung.
- c. Evaluasi algoritma akan menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-score* untuk membandingkan performa masing-masing.

1.3 Rumusan Masalah

Berdasarkan masalah penelitian diatas solusi yang diusulkan, maka ditetapkan rumusan masalah sebagai berikut. “Bagaimana performa algoritma *Random Forest* dan *Decision Tree* dalam memprediksi resiko terkena penyakit jantung berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score*, serta menentukan algoritma mana yang memiliki performa terbaik dalam memprediksi resiko terkena penyakit jantung?’.

1.4 Tujuan Penelitian

Tujuan pada penelitian ini adalah sebagai berikut:

- a. Menganalisis performa algoritma *Random Forest* dan *Decision Tree* dalam memprediksi resiko terkena penyakit jantung berdasarkan metrik berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score*.
- b. Membandingkan kedua algoritma tersebut untuk menentukan algoritma mana dengan performa terbaik dalam memprediksi resiko terkena penyakit jantung.
- c. Menyajikan hasil prediksi dari model *Random Forest* dan *Decision Tree* yang dapat dipahami oleh penulis.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a. Memberikan kontribusi dalam bidang ilmu pengetahuan khususnya dengan menambah wawasan tentang efektivitas berbagai algoritma dalam memprediksi resiko tentang kesehatan, khususnya prediksi penyakit jantung.
- b. Memberikan rekomendasi algoritma yang paling efektif untuk mendukung upaya penanganan penyakit jantung.
- c. Meningkatkan kesadaran akan pentingnya deteksi dini penyakit jantung dan mendorong pola hidup sehat serta meningkatkan efisiensi proses deteksi dini kasus prediksi resiko terkena penyakit jantung secara lebih efisien dan cepat.

1.6 Sistematika Penulisan

Agar lebih jelas dan mudah dipahami, maka penulis akan memberikan prosedur dan gambaran dalam menulis skripsi, sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan masalah, manfaat penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini menjelaskan mengenai tinjauan penelitian terdahulu dan teori dasar yang digunakan yang berkaitan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan-tahapan penelitian, data yang digunakan, Teknik pengumpulan data dan metode atau pemodelan yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tahapan, proses dan hasil Perbandingan Algoritma *Random Forest* dan *Decision Tree* Dalam Memprediksi Resiko Terkena Penyakit Jantung.

BAB V PENUTUP

Bab ini berisikan tentang simpulan dari seluruh pembahasan dan saran yang diperlukan untuk perbaikan dimasa yang akan datang.

DAFTAR PUSTAKA

LAMPIRAN