

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penyakit Jantung merupakan salah satu Penyakit Tidak Menular (PTM) yang banyak menyebabkan kematian. Lebih dari 36 juta orang meninggal karena PTM dan lebih dari 9 juta kematian yang disebabkan oleh PTM terjadi sebelum usia 60 tahun. Yang mana 90% dari kematian “dini” tersebut terjadi di negara berpenghasilan rendah dan menengah. Secara global PTM penyebab kematian nomor satu setiap tahunnya adalah penyakit *kardiovaskuler*. Yang mana merupakan penyakit yang di sebabkan oleh gangguan fungsi jantung dan pembuluh darah, seperti: Penyakit Jantung Koroner, Penyakit Gagal Jantung atau payah jantung, Hipertensi dan Stroke. Penderita penyakit jantung koroner, gagal jantung dan stroke banyak ditemukan pada kelompok umur 45-54 tahun, 55-64 tahun dan 65-74 tahun. Namun demikian, berdasarkan diagnosis atau gejala, penyakit jantung koroner, gagal jantung dan stroke cukup banyak pula ditemukan pada penduduk kelompok umur 15-24 tahun[1].

Penyakit jantung terdiri dari serangkaian gangguan yang mempengaruhi jantung. Ini termasuk masalah pembuluh darah seperti masalah detak jantung yang tidak teratur, otot jantung yang lemah, cacat jantung bawaan, penyakit *kardio vaskular* dan penyakit arteri koroner. Tingkat kematian oleh penyakit pada jantung termasuk kelompok tinggi di dunia. Bidang ilmu kedokteran sangat bergantung pada sarana otomatis berbasis komputer untuk diagnosis yang tepat dan akurat serta tepat waktu. Hal ini mengakibatkan pemeliharaan sejumlah besar data terkait pasien setiap hari. Data yang disimpan dapat digunakan sebagai sumber untuk memprediksi kemungkinan penyakit di masa depan yang membuat teknik penambangan data memainkan peran sentral untuk ekstraksi pengetahuan dan prediksi. Prediksi penyakit jantung adalah salah satu cara untuk melakukan prediksi tersebut[2].

Namun masalah yang timbul dari penyakit jantung adalah sulitnya deteksi awal sehingga penderita sadar akan penyakit jantung ketika sudah parah. Deteksi penyakit jantung dapat dilakukan dengan cara konvensional seperti konsultasi dengan dokter spesialis jantung dan tes laboratorium sehingga biaya yang harus dikeluarkan relatif besar sehingga masyarakat yang ekonominya rendah sulit untuk memeriksakan diri mereka atau masyarakat yang ekonominya tinggi merasa sayang jika mengeluarkan banyak biaya[3].

Kemajuan teknologi telah membuat perkembangan dalam ilmu pengetahuan dan teknologi informasi telah sangat pesat bahkan merambah banyak bidang dan salah satu manfaatnya adalah terciptanya sistem prediksi dunia medis salah satu yang mendapat manfaat dari perkembangan teknologi. Berbagai metode klasifikasi banyak diterapkan untuk melakukan prediksi penyakit. Sistem prediksi sangat berguna dalam memperkirakan sesuatu yang kemungkinan akan terjadi berdasarkan data yang ada sehingga kerugian yang timbul dapat ditanggulangi atau keuntungan yang kemungkinan terjadi dapat dimaksimalkan hasilnya. Kemajuan teknologi data mining dalam sistem kesehatan semakin meningkat. Salah satu teknologi yang ada saat ini adalah *machine learning* yang mampu digunakan untuk mendeteksi penyakit jantung[4].

Ada beberapa metode machine learning yang dapat digunakan untuk klasifikasi. Beberapa algoritma yang dapat digunakan dalam membuat pohon keputusan antara lain *ID3*, *C4.5*, dan *C5.0*. *Algoritma C4.5* merupakan pengembangan dari *Algoritma ID3* dimana kekurangan yang dimiliki oleh algoritma *ID3* dapat ditutupi oleh *Algoritma C4.5*. Kelebihan *Algoritma C4.5* yakni memperoleh pohon keputusan yang mempunyai tingkat akurasi yang baik dan efisien dalam menangani variabel yang bertipe diskrit atau numerik[5].

Pada penelitian sebelumnya penggunaan metode *Algoritma C4.5* untuk mendiagnosis penyakit jantung koroner di dapatkan hasil bahwa *Algoritma C4.5* mudah dipahami dan dapat digambarkan dalam bentuk pohon keputusan berdasarkan faktor resiko yang dialami pasien. Faktor resikonya antara lain umur, jenis kelamin, faktor keturunan, tekanan darah sistolik, tekanan darah diastolik,

kadar kolesterol total, kadar *HDL (High Density Lipoprotein)*, kadar *LDL (Low Density Lipoprotein)* dan kadar *trigliserida*. Berdasarkan pengujian menggunakan *confusion matrix* dengan parameter tingkat akurasi diperoleh nilai sebesar 94,4%[6].

Oleh karena itu peneliti berharap sistem informasi dengan menggunakan metode *Algoritma C4.5* mampu menyelesaikan permasalahan mengenai identifikasi penyakit jantung yang sudah dijelaskan diatas agar mendiagnosa suatu jenis penyakit jantung menjadi lebih efektif dan efisien.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Dalam mendiagnosa suatu penyakit sering berfokus pada naluri dan pengalaman dokter, namun bukan bersumber pada data yang dicakup di dalam *database*. Setelah data terkumpul, perlu dilakukan analisis yang cermat untuk mendeteksi penyakit jantung. Metode-metode seperti *machine learning* dan *data mining* dapat digunakan untuk menganalisis data secara efektif. Teknik ini dapat membantu mengidentifikasi pola dan tren yang berkaitan dengan penyakit jantung, serta memprediksi risiko penyakit jantung pada individu berdasarkan faktor risiko yang ada.

1.2.2 Metode Usulan

Metode usulan yang dapat menyelesaikan masalah yang disebutkan pada bagian 1.2.1 adalah metode *Algoritma C4.5* untuk mengidentifikasi penyakit sejak dini.

1.2.3 Pertanyaan Penelitian

1. Apakah *Algoritma C4.5* efektif dalam mendeteksi dini penyakit jantung berdasarkan data faktor resiko atau atribut yang ada?
2. Bagaimana pengaruh jumlah dataset terhadap performa aplikasi (akurasi)?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan Penelitian

1. Mengetahui seberapa efektif penggunaan *Algoritma C4.5* dalam mendeteksi penyakit jantung secara dini.
2. Dapat mengetahui pengaruh dataset terhadap performa aplikasi (akurasi).

1.3.2 Manfaat

Manfaat dari penelitian ini adalah *Algoritma C4.5* memungkinkan pengembangan model prediksi yang mudah dipahami dan dapat digunakan secara praktis oleh tenaga medis. Model prediksi yang dibangun dengan menggunakan *Algoritma C4.5* dapat memberikan pedoman dalam mengidentifikasi individu dengan risiko tinggi terkena penyakit jantung. Hal ini dapat membantu dalam pengambilan keputusan klinis yang tepat, serta memberikan dukungan dalam perencanaan perawatan dan manajemen penyakit.

1.4 Batasan-Batasan

Adapun batasan-batasan masalah sebagai berikut:

- 1) Data yang digunakan pada penelitian kali ini adalah data *Heart Disease* dari situs *Kaggle Repository*:
(<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>). Dimana data di akses pada 26 Juli 2022.
- 2) Banyak datanya yaitu 1025 baris dengan 13 fitur.

1.5 Sistematika Penulisan Proposal

Penulisan pengerjaan Tugas Akhir ini berdasarkan dari beberapa pengumpulan dengan cara:

(1) BAB I PENDAHULUAN

Bab ini berisikan latar belakang, perumusan masalah, tujuan dan manfaat, batasa-batasan, dan sistematika penulisan proposal.

(2) BAB II KAJIAN PUSTAKA

Pada bab ini terdapat penjelasan mengenai pengertian klasifikasi, penyakit jantung, *algoritma C4.5* dan penelitian terkait.

(3) BAB III METODE USULAN

Bagian ini berisikan tahapan penelitian, arsitektur sistem, flowchart, dan dataset.

(4) BAB IV HASIL DAN PEMBAHASAN

Membahas hasil dari implementasi mengolah data pertama kali, implementasi program, dan hasil dari aplikasi ketika sudah dijalankan. Beserta visualisasi dan data baru yang akan di prediksi.

(5) BAB V PENUTUP

Berisi tentang kesimpulan dan saran dari keseluruhan penelitian yang sudah di kerjakan.