

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Berdasarkan tingkat era globalisasi penyakit jantung merupakan sebutan yang sering digunakan untuk menggambarkan gangguan terhadap fungsi kerja jantung. Adapun jumlah kematian akibat penyakit jantung ini secara global mencapai 18,6 juta orang setiap tahunnya. Penyakit jantung adalah salah satu penyakit yang memiliki angka kematian tinggi. Menurut WHO (Organisasi Kesehatan Dunia) dan CDC, penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada dan Australia. Jumlah orang dewasa yang didiagnosis dengan penyakit jantung terdiri dari 26,6 Juta Jiwa (11,3%) dari populasi orang dewasa [1], maka dari itu diperlukan deteksi dini apakah seseorang terindikasi atau tidaknya mengidap penyakit jantung. Karena jika terlambat ditangani, hal ini dapat berujung pada kematian.

Seiring berkembangnya ilmu pengetahuan, kasus tentang prediksi penderita penyakit jantung dapat diselesaikan menggunakan teknik data mining. Data mining memiliki beberapa teknik diantaranya prediktif atau klasifikasi, klustering dan asosiasi. Dalam proses diagnosa masalah yang sering kali terjadi adalah kurangnya akurasi pada proses klasifikasi. Untuk mengukur tingkat akurasi pada dataset dapat dilakukan dengan teknik klasifikasi. Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia [2].

Klasifikasi merupakan bagian dari pada teknik klasifikasi data mining. Algoritma C5.0 memiliki kemampuan untuk mengklasifikasikan sebagai pohon keputusan atau seperangkat aturan, karena merupakan *classifier* yang mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan *classifier* lain [3]. Jadi, perbedaan yang dapat dilihat secara fisik yaitu *tree* yang akan dihasilkan oleh algoritma C5.0 akan lebih ringkas jika dibandingkan dengan algoritma C4.5. Jika pada algoritma C4.5 berhenti sampai perhitungan *information gain*, maka pada algoritma C5.0 akan melanjutkannya dengan perhitungan *gain ratio* dengan menggunakan *information gain* dan *entropy* yang telah ada. Karena itulah pembangunan *tree* pada algoritma C5.0 lebih cepat dibandingkan algoritma C4.5.

Berdasarkan penjelasan yang sudah dijelaskan diatas, maka penulis akan melakukan penelitian ini yaitu untuk klasifikasi pada dataset penyakit jantung yang diperoleh dari situs Kaggle dengan mengetahui kombinasi setiap atribut pada dataset dan kemudian melakukan perhitungan untuk mengetahui tingkat akurasi pada dataset menggunakan teknik klasifikasi data mining yang berjudul **“IMPLEMENTASI ALGORITMA C5.0 UNTUK KLASIFIKASI DATA PENYAKIT JANTUNG”**

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah yang telah diuraikan diatas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana mengklasifikasi dan menganalisis dataset penderita penyakit jantung menggunakan algoritma C5.0
2. Bagaimana mengevaluasi hasil tingkat akurasi dari prediksi data penderita penyakit jantung dengan menggunakan algoritma C5.0

1.3 BATASAN MASALAH

Agar penelitian ini dapat lebih terarah dan tidak menyimpang dari permasalahan, maka penulis menetapkan batasan masalah sebagai berikut :

1. Penelitian ini hanya mengklasifikasi dan menganalisis dataset penderita penyakit jantung.
2. Dataset penderita penyakit jantung yang diinputkan diperoleh dari situs kaggle :

<https://www.kaggle.com/datasets/firdaus9914/penyakit-jantung> [4]

3. Algoritma yang digunakan yaitu algoritma C5.0.
4. Pengujian hasil analisis menggunakan *tools* Rapidminer.

1.4 TUJUAN PENELITIAN

Berdasarkan pada permasalahan yang telah disampaikan sebelumnya, penelitian ini memiliki tujuan sebagai berikut :

1. Dapat mengklasifikasi dan menganalisis dataset penderita penyakit jantung.
2. Mendapatkan tingkat akurasi yang lebih baik pada proses klasifikasi dataset penderita penyakit jantung.

1.5 MANFAAT PENELITIAN

Pada penelitian ini diharapkan dapat memberikan manfaat antara lain :

1. Dapat mengetahui tingkat akurasi penderita penyakit jantung.
2. Dapat menggunakan hasil analisis untuk mencegah dan mengetahui variabel penyebab penderita penyakit jantung.
3. Dapat menambah pengetahuan bagi pembaca yang bisa digunakan sebagai sumber informasi atau referensi bagi penelitian lebih lanjut mengenai topik yang sama.

1.6 SISTEMATIKA PENULISAN

Gambaran mengenai hal-hal yang akan dibahas penelitian ini terdiri dari beberapa bab, yaitu:

BAB I : PENDAHULUAN

Pada bab ini menjelaskan secara umum mengenai latar belakang masalah, rumusan masalah dengan batasan-batasan masalah yang digunakan, tujuan, dan manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini berisikan landasan teori yang mencakup teori yang berhubungan dengan pokok permasalahan. Teori-teori ini

mengenai *data mining*, algoritma C5.0, dan penjelasan *tools* yang digunakan untuk menganalisis dalam pembuatan penelitian ini.

BAB III : METODOLOGI PENELITIAN

Pada bab ini berisi tahapan proses yang dilakukan selama mengerjakan penelitian, metode pengumpulan data serta *tools* yang digunakan dalam penelitian ini, baik *hardware* maupun *software*.

BAB IV : ANALISIS DAN HASIL

Pada bab ini menjelaskan tentang gambaran umum objek penelitian, serta menganalisis dan menghitung data dengan menggunakan algoritma C5.0, serta implementasi dan pengujian mengenai hasil dari implementasi perangkat lunak yang telah selesai, pengujian perangkat lunak dan analisa hasil yang dicapai.

BAB V : KESIMPULAN

Bab ini merupakan bab yang mencakup tentang kesimpulan yang berisi keseluruhan uraian yang telah dibahas pada bab-bab sebelumnya, serta saran-saran yang diperlukan.