

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG MASALAH**

Mengingat pesatnya perkembangan teknologi, informasi yang akurat sangat penting untuk kehidupan kita sekarang dan masa depan. *Data Mining* adalah proses menghapus atau memfilter sejumlah besar data melalui serangkaian langkah untuk mengekstraksi informasi yang berguna dari data tersebut. [1]. *Data mining* dapat diterapkan di bidang kesehatan. Keputusan dan prediksi mengenai penyakit yang menjadi sasaran akan dihasilkan dari pengambilan keputusan berdasarkan data dan informasi yang akurat [2].

Ada banyak data kesehatan di industri kesehatan, tetapi sebagian besar belum diproses untuk mengetahui informasi tersembunyi yang dibutuhkan dokter untuk membuat keputusan yang baik. Keputusan dan prediksi penyakit yang tepat sasaran akan dihasilkan dari pengambilan keputusan berdasarkan data dan informasi yang akurat.

Penyakit jantung dapat terjadi karena adanya Kerusakan pada sel otot jantung yang bertugas memompa darah ke seluruh tubuh, yang disebabkan kekurangan oksigen yang dibawa darah ke pembuluh darah di jantung, menyebabkan kegagalan organ jantung dalam memompa darah membuat jantung tidak dapat melaksanakan fungsinya dengan baik [3]. Penyakit jantung memiliki angka kematian yang tinggi dan merupakan penyakit yang ditakuti masyarakat.

Oleh karena itu, klasifikasi data mining diperlukan untuk membantu dokter dalam pengambilan keputusan yang cepat dan akurat.

Banyak penelitian yang memprediksi penyakit jantung dengan teknik klasifikasi *data mining*, diantaranya dilakukan oleh [4] metode yang digunakan yaitu *Algoritma Back Propagation Neural Network* dengan Metode *Adaboost*. Hasil penelitian dengan nilai akurasi hasil prediksi menggunakan *algoritma BPNN* adalah 96,65 % dan *algoritma BPNN* dengan metode *Adaboost* menjadi 99,29 %. Pada penelitian [5] metode yang digunakan yaitu *Extreme Learning Machine (ELM)* dan *Algoritma C4.5*. *Algoritma C4.5* dengan tingkat akurasi yang lebih tinggi dibandingkan *ELM* dikarenakan *Algoritma C4.5* dijabarkan berdasarkan nilai gain tertinggi yang dijadikan nodes dan hasil akhirnya membentuk pohon keputusan.

Berdasarkan latar belakang diatas, maka judul yang diambil dalam Tugas Akhir ini adalah “***Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung***”.

## **1.2 RUMUSAN MASALAH**

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *Algoritma C4.5* dalam mengklasifikasi penyakit gagal jantung?
2. Bagaimana hasil pohon keputusan yang didapat *Algoritma C4.5* dalam mengklasifikasi penyakit gagal jantung?

3. Seberapa besar tingkat akurasi *Algoritma C4.5* apabila digunakan untuk mengklasifikasi penyakit gagal jantung?

### 1.3 BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu:

1. Data yang diperlukan adalah dataset Penyakit gagal Jantung (data penyakit gagal jantung yang telah didapat berjumlah 918 data), dataset diambil dari *website kaggle.com* yang dapat didownload dengan link: (<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>).
2. Variabel yang digunakan dalam penelitian ini adalah *Age*, *Sex*, *ChestPainType*, *RestingBP*, *Cholesterol*, *FastingBP*, *RestingECG*, *MaxHR*, *ExerciseAngina*, *OldPeak*, *ST\_Slope*, dan *HeartDisease* yang memberikan informasi mengenai penyakit gagal jantung.
3. *Tools* (alat bantu) yang digunakan untuk klasifikasi adalah *WEKA* dan *RapidMiner*.

### 1.4 TUJUAN DAN MANFAAT PENELITIAN

#### 1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Menerapkan *Algoritma C4.5* untuk mengklasifikasi penyakit gagal jantung maupun bukan penyakit gagal jantung.
2. Menghasilkan pohon keputusan untuk mempermudah dalam menentukan

penyakit gagal jantung.

3. Mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi Penyakit gagal jantung.

#### **1.4.2 Manfaat Penelitian**

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Dapat memberikan kemudahan dalam mengenali penyakit gagal jantung.
2. Mendapatkan pohon keputusan yang memprmudah dalam klasifikasi penyakit gagal jantung.
3. Mendapatkan akurasi yang terbaik dalam klasifikasi penyakit gagal jantung dengan menggunakan tools *WEKA* dan *RapidMiner* sebagai alat bantu klasifikasi.
4. Menambah wawasan bagi penulis dalam pengklasifikasian *dataset* Penyakit Gagal Jantung.

#### **1.5 SISTEMATIKA PENULISAN**

Sistematika dibuat untuk memberi suatu gambaran yang jelas terhadap penulisan ilmiah agar menghindari terjadinya pembahasan diluar dari tema dan judul peneltian. Secara garis besar penulisan ilmiah terdiri dari 5 bab, dapat dilihat dari sistematika penulisan sebagai berikut :

**BAB I : PENDAHULUAN**

Bab ini membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

**BAB II : LANDASAN TEORI**

Bab ini membahas tentang teori-teori dan pendapat para ahli yang berhubungan dengan permasalahan yang dianalisis. Teori-teori yang digunakan antara lain mengenai data mining, klasifikasi, penyakit gagal jantung.

**BAB III : METODOLOGI PENELITIAN**

Bab ini membahas secara rinci kerangka kerja penelitian, metode atau pendekatan yang digunakan, dan *tools* (alat bantu) yang digunakan dalam menjawab masalah penelitian untuk mencapai suatu tujuan.

**BAB IV : ANALISA**

Pada bab ini dilakukan perhitungan dan analisis menggunakan metode *Decision Tree* dengan *Algoritma C4.5* terhadap dataset penyakit gagal jantung. Kemudian akan ditampilkan hasil dari analisis dalam bentuk visualisasi dari *tools WEKA* dan *RapidMiner* yang digunakan.

**BAB V : KESIMPULAN**

Bab ini berisi kesimpulan dan saran yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil yang telah dilakukan.