

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG MASALAH**

Dalam era yang terus berubah dengan cepat, terutama dalam perkembangan teknologi yang terus maju, aliran informasi mengalami peningkatan harian yang signifikan. Peningkatan ini telah menyebabkan akumulasi besar data, baik secara online maupun offline. Bertambahnya jumlah dokumen yang terkumpul setiap hari menciptakan tantangan dalam mengidentifikasi dokumen yang relevan sesuai dengan kebutuhan yang ada. Dengan jumlah data yang melimpah karena kemajuan teknologi informasi yang telah merambah ke hampir semua aspek kehidupan, ada kebutuhan mendesak untuk mengakses dan menggunakan informasi serta pengetahuan yang tersimpan dalam data tersebut. Oleh karena itu, muncul konsep Data Mining, yang bertujuan untuk mengungkap pengetahuan dan informasi berharga yang mungkin tersembunyi dalam basis data besar atau volume data yang sangat besar. Data Mining melibatkan beragam teknik yang digunakan untuk menghasilkan wawasan yang bermanfaat dari dalam basis data besar atau volume data yang sangat besar [1][2].

Gagal jantung merupakan salah satu faktor utama yang berkontribusi terhadap tingginya tingkat morbiditas dan mortalitas di seluruh dunia [3]. Kondisi gagal jantung merujuk pada ketidakmampuan jantung dalam memompa darah dengan efisien untuk memenuhi kebutuhan tubuh akan oksigen dan nutrisi [4]. Dalam dunia medis, gagal jantung seringkali sulit untuk didiagnosis secara klinis,

karena gejala dan tanda klinisnya sangat bervariasi, dan beberapa di antaranya tidak spesifik pada tahap awal penyakit [5]. Mengingat prevalensi yang tinggi dari pasien dengan gagal jantung dan pentingnya organ vital seperti jantung, prediksi dan deteksi dini kejadian gagal jantung menjadi prioritas penting bagi praktisi kesehatan. Namun, hingga saat ini, kemampuan untuk memprediksi kejadian gagal jantung dalam praktik klinis masih seringkali belum mencapai tingkat akurasi yang diharapkan. Oleh karena itu, perlu dikembangkan pendekatan yang canggih dan tepat dalam pengolahan data yang dapat memberikan dasar untuk merumuskan rencana perawatan yang mampu memprioritaskan upaya pengelolaan penyakit, dengan tujuan mengurangi tingkat rawat inap dan tingkat kematian pada pasien yang menderita gagal jantung.

Salah satu cara yang dapat digunakan untuk mengatasi permasalahan tersebut adalah dengan menggunakan teknik *data mining*. *Data mining* merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui [6]. Sejumlah metode data mining telah diterapkan dalam konteks medis, terutama dalam hal pengklasifikasian dan prediksi data. Salah satu algoritma pengklasifikasian yang digunakan dalam data mining adalah *Naïve Bayes*. *Naïve bayes* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [7]. *Naïve Bayes* adalah suatu pendekatan pengklasifikasian probabilitas yang simpel, yang melakukan perhitungan sejumlah probabilitas dengan mempertimbangkan frekuensi dan kombinasi nilai dari dataset yang disediakan. Algoritma *Naïve Bayes* memanfaatkan perhitungan probabilitas matematis untuk menentukan probabilitas

keputusan yang benar, berdasarkan informasi tentang objek yang bersangkutan. Namun, untuk meningkatkan tingkat klasifikasi, baik dalam hal akurasi maupun nilai AUC, diperlukan optimasi pada algoritma *Naïve Bayes*. Salah satu metode optimasi yang digunakan adalah Seleksi Fitur *Greedy Forward Selection*, yang bertujuan untuk mengidentifikasi fitur atau atribut yang relevan. Mengenai metode seleksi fitur *greedy*, dimana *greedy* mencari atribut terbaik atau terburuk dari suatu data dengan menggunakan pengukuran *information gain*[8]. Setelah fitur-fitur yang relevan telah diidentifikasi, data dan catatan akan diolah menggunakan algoritma *Naïve Bayes* untuk menghasilkan hasil kinerja dari model yang telah dibuat.

Berdasarkan latar belakang yang telah dibahas, penulis memilih untuk menggunakan algoritma *Naïve Bayes* serta metode *Greedy Forward Selection* dalam upaya memprediksi penyakit gagal jantung. Hasilnya, penulis Menyusun judul penelitian ini menjadi “**PENERAPAN ALGORITMA NAÏVE BAYES DAN GREEDY FORWARD SELECTION UNTUK PENYAKIT GAGAL JANTUNG**”

## **1.2 PERUMUSAN MASALAH**

Berdasarkan latar belakang yang sudah dijelaskan, maka dapat diambil suatu rumusan masalah yaitu sebahai berikut :

1. Bagaimana menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi penyakit gagal jantung dengan *tools RapidMiner* dan *Weka*.

2. Berapa tingkat akurasi dari penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi penyakit gagal jantung dengan *tools RapidMiner* dan *Weka*.

### 1.3 BATASAN MASALAH

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi :

1. Pada penelitian ini peneliti menggunakan klasifikasi dengan metode *Naïve Bayes* dan *Greedy Forward Selection*.
2. Dataset ini menggunakan data sekunder dari website *Kaggle.com*.
3. Atribut yang digunakan yaitu : *Age, Sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal* dan *target*.
4. Alat bantu (*tools*) yang digunakan pada penelitian ini adalah *RapidMiner* dan *Weka*.

### 1.4 TUJUAN DAN MANFAAT PENELITIAN

#### 1.4.1 Tujuan Penelitian

Berdasarkan permasalahan yang tertera diatas, maka dapat disimpulkan bahwa tujuan dari penelitian ini adalah :

1. Menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi penyakit gagal jantung
2. Mengetahui tingkat akurasi dari penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection*

### 1.4.2 Manfaat Penelitian

Serta manfaat yang akan didapat dari penelitian, yaitu:

1. Dapat melihat bagaimana pengaruh dari atribut yang memiliki banyak data dalam memprediksi penyakit gagal jantung
2. Penulis dapat menambah ilmu dan wawasan baru mengenai penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi penyakit gagal jantung
3. Dapat digunakan sebagai acuan untuk melakukan penelitian selanjutnya.

### 1.5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan tugas akhir ini, penulis menguraikan dalam beberapa bab yaitu :

#### **BAB I : PENDAHULUAN**

Pada bab ini akan diuraikan tentang latar belakang masalah, rumusan masalah, matasan masalah, tujuan dan manfaat penulisan, dan sistematika penulisan

#### **BAB II : LANDASAN TEORI**

Pada bab landasan teori 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, prediksi, penyakit gagal jantung, *Naïve Bayes*, *Greedy Forward Selection*, dan alat bantu aplikasi *RapidMiner* dan *weka*.

**BAB III : METODOLOGI PENELITIAN**

Pada bab ini menjelaskan tentang penjelasan dari metode-metode penelitian yang digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian

**BAB IV : ANALISIS**

Pada bab ini menjelaskan tentang gambaran umum objek penelitian, serta menganalisis dataset yang didapat dari website Kaggle.com dengan melakukan persiapan data, kemudian penulis akan melakukan *Exploratory Data Analysis* untuk mengetahui prediksi tingkat akurasi data menggunakan metode klasifikasi algoritma *Naïve Bayes* dan menggunakan seleksi fitur *Greedy Forward Selection* serta menggunakan alat bantu atau *tools* yaitu *RapidMiner*. Hasil dari analisis yang telah dilakukan diinterpretasikan menjadi informasi mengenai status prediksi penyakit gagal jantung serta menggunakan alat bantu *tools* yaitu *RapidMiner*.

**BAB V : KESIMPULAN**

Bab ini berisikan kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.