

# BAB I

## PENDAHULUAN

### 1.1 LATAR BELAKANG MASALAH

*Data mining* berperan sangat penting dalam mengidentifikasi pola tersembunyi dalam data besar. *Data mining* merupakan upaya untuk menggali pengetahuan dan informasi bermanfaat dari berbagai sumber, seperti basis data atau informasi yang tersimpan dalam *database* [1]. *Data mining* merupakan sebuah proses dari *Knowledge Discovery in Databases (KDD)*, dimana pada teknik *data mining* berusaha mendapatkan informasi yang penting dari *database* [2]. *Data mining* tidak hanya menjadi alat analisis, melainkan juga menjadi sarana yang memperkaya wawasan untuk meningkatkan pemahaman dan pengambilan keputusan berbasis data.

Klasifikasi memiliki tujuan untuk membangun model yang dapat memetakan item data ke kategori tertentu, berdasarkan data yang ada [3]. Model klasifikasi merupakan teknik memprediksi data, membuat prediksi nilai dari suatu data yang hasilnya telah ditemukan berasal dari data yang berbeda [4]. Klasifikasi membantu dalam membuat model yang dapat mengkategorikan data ke dalam kelas-kelas yang telah ditentukan sebelumnya, meningkatkan pemahaman tentang pola atau hubungan di dalam *dataset*.

*Naive Bayes* merupakan metode dalam bidang *data mining* yang memanfaatkan perhitungan probabilitas dan statistik yang melakukan prediksi di masa yang akan datang berdasarkan pengalaman sebelumnya [5]. *Naive Bayes*

merupakan salah satu algoritma yang efisien untuk melakukan klasifikasi pada dataset yang besar [6]. *Naive Bayes* mampu membuat prediksi yang tepat dengan memperhitungkan probabilitas dan hubungan statistik antara variabel-variabel yang ada. Ini menjadikan *Naive Bayes* alat yang efektif untuk klasifikasi dan prediksi dalam berbagai konteks, memberikan solusi yang handal berdasarkan pola yang teridentifikasi dari pengalaman sebelumnya.

Industri kesehatan memiliki sejumlah besar data yang belum dimanfaatkan sepenuhnya untuk mengungkap informasi rahasia yang penting bagi dokter dalam membuat keputusan yang optimal. Keputusan dan prediksi penyakit yang tepat dapat diperoleh dengan memanfaatkan data dan informasi yang akurat seperti penyakit gagal jantung.

Penyakit gagal jantung umumnya terjadi karena adanya kerusakan pada sel-sel otot jantung yang mengakibatkan gangguan dalam memompa aliran darah ke seluruh tubuh. Ini disebabkan oleh kurangnya pasokan oksigen ke pembuluh darah jantung atau bahkan karena terjadi spasme pada otot jantung, yang mengakibatkan kegagalan organ jantung dalam memompa darah. Akibatnya, kondisi jantung tidak dapat menjalankan fungsinya dengan baik [7]. Penyakit gagal jantung memiliki angka kematian yang tinggi dan menciptakan rasa takut di kalangan masyarakat. Oleh karena itu, penggunaan klasifikasi dalam *data mining* menjadi penting untuk mendukung dokter dalam pengambilan keputusan yang akurat dan sesuai.

*Dataset* memiliki peranan penting dalam konteks klasifikasi *data mining* karena data ini berperan membantu para peneliti dan praktisi dalam mengembangkan model yang mampu melakukan prediksi penyakit jantung. *Dataset* yang baik untuk digunakan dalam pengklasifikasian merupakan *dataset* yang tersedia di *Kaggle*. *Dataset* dengan nama *Heart Failure Prediction*. *Dataset* ini memiliki beberapa atribut yaitu *Age*, *Sex*, *ChestPainType*, *RestingBP*, *Cholesterol*, *FastingBP*, *RestingECG*, *MaxHR*, *ExerciseAngina*, *Oldpeak*, *ST\_Slope*, dan *HeartDisease*. Selain itu, *dataset* ini juga memiliki jumlah data yang cukup banyak, yaitu 918 data.

Penelitian yang akan dilakukan terhadap pengklasifikasian penyakit jantung sebelumnya sudah pernah dilakukan oleh beberapa peneliti, diantaranya seperti penelitian yang dilakukan oleh Cahya Putri Buani [8], menghasilkan bahwa metode *Naive Bayes* pada prediksi gagal jantung hanya mencapai akurasi sebesar 69,60%, tetapi setelah melalui seleksi *feature* dengan menggunakan algoritma genetika, akurasi meningkat menjadi 96,67%.

Penelitian sejenis lainnya yang telah dilakukan oleh Sari dkk. [9], terhadap *dataset* UCI mengenai klasifikasi penyakit jantung menggunakan metode *Neural Network* memiliki kemampuan dalam meningkatkan akurasi pada Algoritma *Naive Bayes* untuk menentukan klasifikasi resiko Penyakit Jantung dengan hasil nilai akurasi terbaik sebesar 85,33% *presisi* sebesar 94,84% dan *recall* sebesar 88,55% menggunakan data dari UCI *dataset*. Metode *Neural Network* yang mendapatkan hasil lebih rendah yaitu 78,55%.

Berdasarkan permasalahan diatas sangat melatar belakangi Tugas Akhir yang diambil dengan judul **“Penerapan Algoritma *Naive Bayes* Untuk Klasifikasi Penyakit Jantung”**.

## 1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah yang dapat diambil dalam penelitian ini sebagai berikut:

1. Bagaimana menerapkan algoritma *Naive Bayes* untuk klasifikasi penyakit Gagal Jantung pada *dataset Heart Failure Prediction*?
2. Seberapa akurasi algoritma *Naive Bayes* dalam memprediksi penyakit Gagal Jantung pada *dataset Heart Failure Prediction*?

## 1.3 BATASAN MASALAH

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

1. Data yang digunakan adalah *dataset* Penyakit Gagal Jantung yang diambil dari *repository* Kaggle dengan judul *dataset Heart Failure Prediction*. *Dataset* ini terdiri dari 918 data dengan 11 *feature* dan 1 label, sehingga secara keseluruhan terdapat 12 atribut.
2. Algoritma klasifikasi yang digunakan pada penelitian ini yaitu *Naive Bayes*.
3. Alat Bantu yang digunakan untuk klasifikasi adalah *RapidMiner*.

## **1.4 TUJUAN DAN MANFAAT PENELITIAN**

### **1.4.1 Tujuan Penelitian**

Tujuan dari penelitian ini melibatkan beberapa aspek penting, yaitu:

1. Untuk menerapkan algoritma *Naive Bayes* dalam proses klasifikasi Penyakit Gagal Jantung menggunakan *dataset Heart Failure Prediction*.
2. Untuk mengetahui tingkat akurasi algoritma *Naive Bayes* dalam melakukan prediksi Penyakit Gagal Jantung menggunakan *dataset Heart Failure Prediction*.

### **1.4.2 Manfaat Penelitian**

Manfaat dari penelitian ini mencakup berbagai hal, seperti:

1. Penelitian ini dapat menghasilkan model algoritma *Naive Bayes* yang dapat digunakan untuk meningkatkan akurasi klasifikasi Penyakit Gagal Jantung.
2. Mendapatkan tingkat akurasi yang terbaik dalam mengklasifikasikan Penyakit Gagal Jantung dengan algoritma *Naive Bayes*.
3. Penelitian ini dapat menjadi acuan bagi peneliti selanjutnya dalam meningkatkan akurasi untuk klasifikasi Penyakit Gagal Jantung.
4. Penulis mendapatkan wawasan baru dalam hal pengklasifikasian pada *dataset* Penyakit Gagal Jantung.

## 1.5 SISTEMATIKA PENELITIAN

Sistematika dari penelitian ini adalah:

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

Pada bab ini, penulis menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian yang menjadi dasar dari dilaksanakannya penelitian ini. Dan pada bab ini penulis menjelaskan hasil dari penelitian yang telah dilakukan sebelumnya.

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

Pada bab landasan teori ini berisikan teori-teori yang berasal dari para ahli dan berkaitan dengan permasalahan yang akan dibahas. Adapun teori yang akan dibahas yaitu mengenai *data mining*, klasifikasi, algoritma *naïve bayes*, penyakit gagal jantung, *RapidMiner* dan penelitian sejenis.

### - **BAB III: METODOLOGI PENELITIAN**

Pada bab metodologi penelitian ini untuk menjelaskan tahap-tahap penelitian, maka bab ini akan menjelaskan proses tersebut yaitu, kerangka penelitian, metode pengumpulan data, metode klasifikasi serta alat yang digunakan selama penelitian berlangsung.

- **BAB IV: ANALISIS DAN HASIL**

Pada bab ini akan dilakukan analisis perhitungan dengan algoritma *Naïve Bayes* pada *dataset* penyakit Gagal Jantung serta pengujian akurasi menggunakan *software Rapidminer*.

- **BAB V: PENUTUP**

Pada bab ini, bagian akhir dari penelitian yang akan dibahas. Bab penutup ini akan menjelaskan tentang kesimpulan dari penelitian yang dilakukan serta saran yang akan bermanfaat bagi pihak-pihak yang berpartisipasi dan ikut serta dalam proses penelitian ini.