

# **BAB I**

## **PENDAHULUAN**

### **1.1 LATAR BELAKANG MASALAH**

Saat ini, kita hidup pada zaman modernisasi. Modernisasi sendiri merujuk pada suatu bentuk transisi dari keadaan yang kurang berkembang atau kurang maju ke keadaan yang lebih baik dengan harapan dapat memperbaiki kehidupan. Di era modern saat ini, masyarakat sangat bergantung pada teknologi, perkembangan teknologi yang pesat telah merambah ke berbagai bidang termasuk kesehatan. Dengan bantuan teknologi, dapat membantu dalam banyak pekerjaan dibidang medis. Selain kemampuan menyaring data dan mengubahnya menjadi informasi, juga memiliki kemampuan menyimpan kapasitas yang jauh lebih besar. Konvergensi teknologi informasi juga memudahkan dan mempercepat berbagi data kesehatan.

Gagal jantung merupakan penyebab utama morbiditas dan mortalitas di seluruh dunia. Gagal jantung didefinisikan sebagai ketidakmampuan jantung untuk memompa darah untuk memenuhi kebutuhan tubuh akan oksigen dan nutrisi [1]. Gagal jantung sulit dikenali secara klinis, karena presentasi klinis yang beragam dan beberapa tanda klinis yang tidak spesifik pada tahap awal penyakit. Mengingat tingginya jumlah pasien dengan gagal jantung dan pentingnya organ vital seperti jantung, prediksi gagal jantung telah menjadi prioritas bagi dokter, namun sejauh ini prediksi kejadian gagal jantung pada praktek klinis biasanya belum dapat dicapai dengan akurasi yang tinggi. Oleh karena itu, diperlukan cara yang tepat untuk

mengolah data ini untuk menghasilkan rencana kesehatan yang dapat memprioritaskan upaya pengelolaan penyakit untuk mengurangi rawat inap dan kematian pada pasien gagal jantung.

Salah satu cara yang dapat digunakan untuk mengatasi permasalahan tersebut adalah dengan menggunakan teknik *data mining*. Beberapa teknik *data mining* telah diterapkan pada dunia medis seperti dalam kasus klasifikasi dan data prediktif. Salah satu algoritma klasifikasi dalam *data mining* yang dapat digunakan adalah *Naïve Bayes*. *Naïve Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. *Naïve Bayes* menggunakan perhitungan matematis probabilitas dengan syarat nilai keputusannya benar berdasarkan informasi objek. Namun algoritma *naïve bayes* diperlukan optimasi untuk meningkatkan tingkat klasifikasi baik berdasarkan akurasi atau nilai AUC (*Area Under the Curve*) menggunakan seleksi fitur *Greedy Forward Selection* sehingga akan mendapatkan fitur atau atribut yang relevan. Kemudian fitur dan record akan diproses menggunakan algoritma *naïve bayes* untuk mendapatkan hasil dari kinerja model yang dibuat.

Penelitian ini bukanlah penelitian yang pertama dilakukan, penelitian terkait menggunakan dataset yang sama dengan judul penelitian “CLASSIFICATION OF HEART FAILURE USING THE NAÏVE BAYES ALGORITHM” dengan hasil tingkat akurasi menggunakan metode *naïve bayes* (86.18%), *precision* (87.01%), *recall* (88.16%), And AUC (91.2%)[2]. Lalu, penelitian terkait selanjutnya menggunakan dataset berbeda dengan judul penelitian “*Machine learning can*

*predict survival of patients with heart failure from serum creatinine And ejection fraction alone*” dengan tingkat akurasi menggunakan metode *Random Forests* 74,00% , dengan metode *Decision Tree* 73,70% , *Gradient Boosting* 73,80% , *Linear Regression* 73,00% , *One Rule* 72,90% , *Artificial Neural Network* 68,00% , *Naïve Bayes* 69,60% , *SVM Radial* 69,00% , *SVM linear* 68,40% , dan *K-Nearest Neighbors* 62,40% [3]. Dari beberapa penelitian tersebut bahwa metode *Random Forest* memiliki akurasi yang lebih tinggi dari pada metode yang lain, tingkat akurasinya mencapai 74,00% , sedangkan *naïve bayes* memiliki akurasi yang rendah, oleh karena itu didalam penelitian ini akan dilakukan peningkatan akurasi dari *naïve bayes* dengan menggunakan seleksi fitur *Greedy Forward Selection*.

Berdasarkan latar belakang permasalahan yang telah dibahas, penulis memutuskan untuk menerapkan algoritma *naïve bayes* dan *Greedy Forward Selection* untuk prediksi gagal jantung. Yang Penulis tuangkan dalam judul **“PENERAPAN ALGORITMA NAÏVE BAYES DAN GREEDY FORWARD SELECTION UNTUK PREDIKSI GAGAL JANTUNG”**.

## **1.2 PERUMUSAN MASALAH**

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

1. Bagaimana menerapkan algoritma *naïve bayes* dan *Greedy Forward Selection* untuk prediksi gagal jantung.
2. Berapa tingkat akurasi dari penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi gagal jantung.

### 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 metode klasifikasi dengan metode *Naïve Bayes* dan *Greedy Forward Selection*.
2. Dataset ini menggunakan data sekunder dari website *Kaggle.com* (Fedesoriano, 2021).
3. Atribut yang di gunakan yaitu : *Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, ST\_Slope*, dan *HeartDisease*.
4. Alat bantu (*tools*) yang digunakan pada penelitian ini adalah *RapidMiner* dan *Google Colaboratory*.

### 1.4 TUJUAN & 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 gagal jantung.
2. Mengetahui tingkat akurasi dari penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi gagal jantung.

### 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 gagal jantung.
2. Penulis dapat menambah ilmu dan wawasan baru mengenai menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi 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, batasan masalah, tujuan dan manfaat penulisan, dan sistematika penulisan.

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

Pada Bab IAndasan 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, gagal jantung, *naïve bayes* , *Greedy Forward Selection* dan alat bantu aplikasi *RapidMiner*.

**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 gagal jantung serta menggunakan alat bantu atau *tools* yaitu *RapidMiner*.

**BAB V : KESIMPULAN**

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