

BAB I

PENDAHULUAN

1.1 Latar Belakang

Arrhythmia merupakan salah satu kelainan pada jantung yang ditandai dengan ritme jantung yang tidak normal, menjadi lebih cepat atau lebih lambat. Untuk mendiagnosa *arrhythmia* tidak dapat dilakukan hanya dengan pengamatan fisik karena beberapa tipe *arrhythmia* tidak memiliki gejala yang dirasakan pasien. Untuk mendiagnosa *arrhythmia* perlu dilakukan prosedur *Electrocardiograf*. *Electrocardiograf* adalah proses merekam aktivitas listrik jantung dengan cara menempelkan elektroda pada bagian dada, lengan, dan tungkai pasien. *Electrocardiograf* ini akan mendeteksi perubahan pola *depolarisasi* dan *repolarisasi* setiap detak jantung. Hasil rekaman tersebut berupa *Electrodiogram* atau yang biasa di kenal sebagai ECG. ECG inilah yang kemudian dijadikan rujukan untuk mendiagnosa *arrhythmia*.

Untuk mendiagnosa *arrhythmia* perlu dilakukan proses analisa terhadap data ECG agar didapatkan diagnosa yang sesuai dengan keadaan pasien sesungguhnya. Menganalisa data ECG bukanlah hal mudah. Data ECG yang cukup besar tersebut menjadi hambatan tersendiri dalam proses analisa.

Klasifikasi merupakan salah satu pendekatan yang dapat digunakan untuk menganalisa data ECG. Hasil analisa tersebut dapat dijadikan rujukan untuk mendiagnosa *arrhythmia* berdasarkan hasil dari ECG. Salah satu algoritma klasifikasi yang sering digunakan adalah *Naïve Bayes*. Algoritma ini menggunakan probabilitas dan statistik berdasarkan teorema *Bayes*.

Penelitian ini akan menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan data *arrhythmia*. Pemilihan *Naïve Bayes* berdasarkan beberapa penelitian yang telah dilakukan sebelumnya menggunakan algoritma *Naïve Bayes*. Beberapa kelebihan dari *Naïve Bayes*, antara lain: memiliki performa yang lebih baik dari beberapa algoritma lainnya [1] [2], tidak membutuhkan data yang besar untuk memulai proses pelatihan [3], dan *Naïve Bayes* memiliki tingkat akurasi dan kecepatan yang tinggi ketika di aplikasikan pada data dalam jumlah besar [4]. Maka dari itu, penelitian ini menggunakan *Naïve Bayes* sebagai algoritma klasifikasi dikarenakan performa yang dimilikinya.

Data dengan jumlah besar mengandung banyak fitur, baik fitur yang relevan, tidak relevan, ataupun fitur yang redundan. Membiarkan fitur-fitur yang tidak relevan dan redundan akan membingungkan proses klasifikasi data sehingga akan menurunkan kecepatan klasifikasi, meningkatkan biaya komputasional dan penggunaan memori, serta akan sangat berpengaruh terhadap hasil klasifikasi itu sendiri [4] [5]. Banyaknya fitur dapat menyebabkan *overfit* pada model sehingga menyebabkan penurunan performa yang dimiliki oleh model. Oleh karena itu, perlu dilakukan tahapan *preprocessing* untuk pemilihan fitur yang relevan.

Seleksi fitur merupakan teknik pemilihan fitur yang relevan berdasarkan kriteria tertentu sehingga dapat meningkatkan performa pelatihan seperti meningkatkan akurasi klasifikasi, menurunkan biaya komputasi dan interpretasi model yang lebih baik [6]. Seleksi fitur memiliki tiga model yang biasa digunakan yaitu *wrapper*, *filter*, dan *embedded*. Keunggulan dari seleksi fitur *filter* ialah memiliki waktu komputasi yang lebih rendah dari pada tipe lainnya, sederhana dan cepat, mengukur data dimensi tinggi dengan mudah, serta bersifat independen dari algoritma klasifikasi. Salah satu algoritma seleksi fitur tipe *filter* adalah *Information Gain*. Algoritma ini dapat menangani proses seleksi fitur dengan cepat [7], dan lebih efektif dalam menghilangkan fitur dengan akurasi yang baik [8]. Dari kelebihan yang dipaparkan diatas, maka penelitian ini akan menggunakan seleksi fitur dengan tipe *filter*. Algoritma yang akan digunakan adalah *Information Gain*.

Oleh sebab itu pada penelitian ini akan membahas mengenai pengaruh seleksi fitur terhadap performansi model klasifikasi *Naïve Bayes* pada diagnosa *arrhythmia*. Dengan penerapan seleksi fitur diharapkan dapat meningkatkan performa model klasifikasi *Naïve Bayes* pada diagnosa *arrhythmia*.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas maka didapatkan perumusan masalah sebagai berikut:

1. Bagaimana pengaruh seleksi fitur dengan metode *filter* dalam meningkatkan akurasi klasifikasi *Naïve Bayes* pada diagnosa *arrhythmia*?
2. Berapa akurasi yang dihasilkan dari komparasi model klasifikasi dengan seleksi fitur?

1.3 Tujuan dan Manfaat

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui pengaruh seleksi fitur terhadap performansi pada model klasifikasi *Naïve Bayes* pada diagnosa *arrhythmia* dengan membandingkan hasil klasifikasi *Naïve Bayes* tanpa menerapkan seleksi fitur dengan hasil klasifikasi *Naïve Bayes* dengan penerapan seleksi fitur.

Manfaat dari penelitian ini adalah untuk mengetahui fitur-fitur yang relevan dari sebuah dataset *arrhythmia* setelah penerapan seleksi fitur sehingga dapat meningkatkan performa klasifikasi *Naïve Bayes*.

1.4 Batasan Masalah

Batasan-batasan dalam penelitian ini adalah:

1. Data input berasal dari *Large Dataset, UCI Machine Learning Repository*. Dataset yang digunakan merupakan dataset *arrhythmia*. Dataset ini terdiri dari 452 jumlah data dan 262 fitur. Dataset ini merupakan data dari 452 pasien yang terdiri dari data pribadi pasien seperti jenis kelamin, umur, tingi, berat badan dan data ECG.
2. Dataset yang akan digunakan merupakan dataset yang telah bersih dari *noise* dan *missing value*.
3. Metode seleksi fitur yang akan digunakan adalah tipe *filter* dengan algoritma *Information Gain*

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian Tugas Akhir ini meliputi studi pustaka, pengumpulan data, perancangan dan implementasi sistem, uji coba sistem, analisa dan evaluasi, serta dokumentasi.

1. Studi Pustaka

Mempelajari literatur-literatur yang berkaitan dengan teori dan konsep *arrhythmia*, seleksi fitur, klasifikasi, dan algoritma *Naïve Bayes*.

2. Pengumpulan Data

Tahap berikutnya dalam penelitian ini adalah pengumpulan data. Data yang digunakan dalam penelitian ini diunduh dari *UCI Machine Learning Repository*.

3. Perancangan dan Implementasi Sistem

Sistem dirancang sesuai dengan metode yang diusulkan. Implementasi sistem menggunakan bahasa pemrograman *Python*.

4. Uji Coba Sistem

Menguji model hasil implementasi metode yang telah diusulkan.

5. Analisa dan Evaluasi

Menganalisa dan mengevaluasi hasil uji coba sistem untuk melihat tingkat akurasi dan performa sistem. Evaluasi klasifikasi akan menggunakan metode *Confusion Matrix* dan AUC untuk mengukur tingkat akurasi dari klasifikasi.

6. Dokumentasi

Dokumentasi dilakukan sebagai penyusunan laporan tugas akhir yang berlangsung mulai dari tahap studi pustaka hingga tahap evaluasi serta menarik kesimpulan dari hasil penelitian yang telah dilakukan.

1.6 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian ini, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai *arrhythmia*, seleksi fitur, klasifikasi, dan *cross validation*.

BAB III PERANCANGAN SISTEM

Pada bab ini akan dijelaskan rancangan dari seluruh sistem yang akan dibuat dan alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi sistem dari rancangan yang telah dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan selanjutnya.