

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dalam kehidupan sehari-hari kita tidak terlepas dari sebuah hal kecil yang bernama data. Dari data, kita dapat menemukan suatu informasi dan hal baru yang dapat dipelajari. Namun untuk mengolah data menjadi sebuah informasi dapat dilakukan dengan cara bermacam-macam. Salah satunya dengan menggunakan teknik data mining. Alasan menggunakan data mining, karena data mining adalah salah satu teknik yang digunakan untuk menemukan sebuah informasi penting dari sekumpulan data yang berukuran besar atau berdimensi tinggi, lalu dari data tersebut diubah menjadi sebuah informasi seperti prediksi ataupun sebuah pola yang dapat dimengerti dengan mudah [1].

Data berdimensi tinggi yaitu data yang memiliki puluhan hingga ribuan atribut. Namun jika semakin tinggi dimensi yang terdapat dalam sebuah data, maka kompleksitasnya semakin besar. Sehingga muncul sebuah permasalahan baru yang biasa disebut dengan *the curse of dimensionality* yang menyebabkan semakin banyak biaya dan juga lamanya waktu yang dibutuhkan sebuah algoritma untuk menyelesaikan perhitungan tersebut. Selain itu model perhitungan juga menjadi tidak efektif atau rentan dikarenakan banyaknya *noise* yang ada dalam sebuah data yang berukuran sangat besar [2].

*Evolutionary Algorithm* (EAs) memiliki beberapa fitur yang membuatnya menarik apabila diaplikasikan pada data mining. Salah satunya adalah karena EAs dapat membantu proses pengolahan data dalam mereduksi dimensi/atribut agar didapatkan atribut-atribut optimal dalam mengolah sebuah data. Dengan mereduksi atribut maka kita dapat mengurangi resiko kerentanan sebuah model dan dapat mengatasi masalah optimasinya [3].

*Genetic Algorithm* merupakan salah satu *Evolutionary Algorithm* berbasis evolusi biologi seperti pada dunia nyata. Algoritma ini terus melakukan *cross over* dan mutasi individu hingga didapatkan sebuah individu yang terbaik. Menerapkan konsep *Genetic Algorithm* dalam pengolahan data diharapkan dapat menghasilkan akurasi yang tinggi dan mendapatkan hasil yang optimum.

Banyak penelitian yang telah membahas penerapan *Genetic Algorithm* ke dalam data mining untuk meningkatkan hasil optimasi, diantaranya : Penggunaan *Associative classification* dengan *Genetic Algorithm* untuk memprediksi penyakit jantung [5], Penggunaan *Genetic Algorithm* untuk menentukan *mining rule* pada data

berukuran besar [4], mengoptimasi *Association rule mining* menggunakan *Genetic Algorithm* [8], dan masih banyak lagi. Dari berbagai penelitian yang telah dilakukan, *Genetic Algorithm* selalu dapat meningkatkan akurasi dari sebuah model data mining, terutama dalam kasus klasifikasi.

Maka dari itu dalam tugas akhir ini akan dibahas mengenai analisis prediksi dan bagaimana implementasi *Genetic Algorithm* pada *Associative Classification* dalam memecahkan sebuah permasalahan pada data penyakit yang berukuran sangat besar.

## 1.2 Rumusan Masalah

Adapun rumusan masalah dari tugas akhir ini adalah :

- Bagaimana cara kerja *associative classification* dan *genetic algorithm* dalam menyelesaikan kasus yang memiliki data berdimensi tinggi?
- Berapa besar tingkat akurasi yang didapatkan dari penggabungan *associative classification* dengan *genetic algorithm*?
- Bagaimana membangun model yang mampu memprediksi pasien yang terjangkit penyakit dengan *associative classification* dan *genetic algorithm*?

## 1.3 Tujuan

Adapun tujuan dari tugas akhir ini adalah :

- Mengetahui cara kerja *associative classification* dan *genetic algorithm* dalam menyelesaikan kasus yang memiliki data berdimensi tinggi.
- Mengetahui besar tingkat akurasi yang didapatkan dari penggabungan *associative classification* dengan *genetic algorithm*.
- Membangun model yang mampu memprediksi pasien yang terkena penyakit dengan menggunakan *associative classification* dan *genetic algorithm*

## 1.4 Batasan Masalah

Batasan-batasan masalah yang digunakan dalam tugas akhir ini adalah :

- Menggunakan data dari *Kent Ridge Bio-medical Data Set Repository*. Berupa data penyakit berdimensi tinggi dan bersifat numerik.

## 1.5 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut :

### **BAB I            PENDAHULUAN**

Bab ini terdiri dari latar belakang, perumusan masalah, tujuan, batasan masalah, dan juga sistematika penulisan. Dimana bab ini akan menjelaskan tugas akhir secara umum.

## **BAB II        TINJAUAN PUSTAKA**

Bab ini menjelaskan uraian teori yang berhubungan dengan data berdimensi tinggi, *Associative classification*, dan *Geneic Algorithm (GA)*.

## **BAB III        PERANCANGAN SISTEM**

Pada bab ini berisi rancangan dari sistem yang dibangun dimana pembangunan sistem ini akan menjelaskan langkah-langkah dari proses preprocessing, reduksi atribut, serta pembangunan dua algoritma *Associative classification* dan *Genetic Algorithm (GA)* dalam memprediksi penyakit.

## **BAB IV        PENGUJIAN DAN ANALISIS**

Bab ini berisi analisis hasil pengujian dari system prediksi yang dibangun dengan menggunakan metode *Associative classification* dan *Geneic Algorihm (GA)*.