

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan data beserta pengolahan data selama beberapa tahun ini sangatlah pesat, perkembangan tersebut ditandai dengan semakin pentingnya informasi dan pengolahan data dalam segala jenis aspek kehidupan. Karena dalam kehidupan sehari-hari kita tidak terlepas dari sebuah hal kecil yang bernama data. Definisi Data itu sendiri adalah keterangan yang benar dan nyata [1]. Salah satu data yang sangat berperan penting adalah data penyakit. Data penyakit berperan penting dalam melihat perkembangan penyakit tertentu disuatu daerah. Kemudian data tersebut diolah dan menghasilkan sebuah informasi yang nantinya menjadikan data tersebut sebagai diagnosis dari penyakit.

Untuk mengolah data penyakit tersebut, dibutuhkanlah sebuah metode untuk menggali informasi dari data mentah menjadi data yang memiliki informasi yang sangat bermanfaat, salah satunya adalah *Data mining*. Sebuah penelitian juga menunjukkan bahwa penggunaan data mining dalam uji klinis membantu meringankan biaya [2]. Namun didalam data mining terdapat beberapa tantangan. Misalnya, data dengan dimensi tinggi yang artinya data memiliki puluhan bahkan ribuan atribut. Maka dari itu dibutuhkan sistem yang dapat mereduksi banyaknya atribut. Dari atribut yang sangat banyak ini menimbulkan permasalahan yang sangat kompleks untuk pengolahan sebuah data. Fenomena yang ada dalam data berdimensi tinggi adalah “*Curse of Dimensionality* “ yaitu semakin tinggi dimensi pada sebuah data maka semakin sulit dan semakin menurun nilai informasi yang akan didapatkan [3].

Penggabungan *Data mining* dengan *Evolutionary algorithm* (EAs) atau sering disebut *Evolutionary Data Mining* (EvoDM) mampu menjawab tantangan yang ada pada data berdimensi tinggi. Dengan banyaknya atribut *Evolutionary Data Mining* (EvoDM) ini dapat mereduksi atribut sehingga mengurangi data yang kurang informatif. Algoritma EAs yang dimaksud adalah *Momentum-type Particle Swarm Optimization* (MPSO). Algoritma EAs ini dikombinasi dengan algoritma *K-Means* sebagai salah satu cara untuk menentukan atribut yang memiliki bobot paling optimal dan pusat cluster dari setiap atribut [4]. Pada penelitian sebelumnya kedua algoritma hybrid ini mendapatkan nilai akurasi yang cukup tinggi dibandingkan dengan menggunakan algoritma k-means saja [4,5]. Pada penulisan tugas akhir ini mengacu pada paper “*Development of Evolutionary Data Mining Algorithms and their Applications to Cardiac Disease Diagnosis*”. Paper dipilih karena penggabungan Algoritma MPSO dan *k-means* menghasilkan bobot optimal dan centroid dari setiap atribut. Data yang digunakan adalah data penyakit jantung UCI Machine Learning Repository, dan menghasilkan 90,37% [4].

Maka dari itu, dalam tugas akhir ini dibangun sebuah sistem yang dapat memprediksi penyakit menggunakan algoritma *momentum-type particle swarm optimization* dengan *k-means* (MPSO-KM) pada data penyakit berdimensi tinggi. Adapun data yang digunakan pada tugas akhir ini merupakan data berdimensi tinggi yang diambil dari *Kent Ridge Bio-Medical Data Set Repository*.

1.2 Rumusan Masalah

Pada tugas akhir ini, rumusan masalah yang dibahas adalah:

1. Bagaimana implementasi algoritma *momentum-type particle swarm optimization* dengan *k-means* (MPSO-KM) untuk data berdimensi tinggi?
2. Bagaimana hasil akurasi yang didapat dari algoritma hybrid MPSO-KM?
3. Bagaimana kinerja algoritma *momentum-type particle swarm optimization* dengan *k-means* (MPSO-KM) untuk data berdimensi tinggi?

1.3 Tujuan

Adapun tujuan dari dilakukannya tugas akhir ini adalah:

1. Mengetahui implementasi dari *momentum-type particle swarm optimization* dengan *k-means* (MPSO-KM) untuk data berdimensi tinggi
2. Mengetahui akurasi dari algoritma hybrid MPSO-KM.
3. Mengetahui kinerja dari algoritma *momentum-type particle swarm optimization* dengan *k-means* (MPSO-KM) untuk data berdimensi tinggi.

1.4 Batasan Masalah

Batasan-batasan yang digunakan untuk mempermudah tugas akhir ini adalah:

1. Menggunakan data dari *Kent Ridge Bio-medical Data Set Repository*. Berupa data penyakit berdimensi tinggi.
2. Untuk r_1 dan r_2 adalah bilangan random uniform yang ada pada interval $[0,1]$ yaitu 0.1, 0.3, 0.5, 0.7, 0.9.

1.5 Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika penulisan sebagai berikut:

BAB I Pendahuluan

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

BAB II Kajian Pustaka

Bab ini menjelaskan dasar teori yang berhubungan dengan data berdimensi tinggi, Momentum of-type Particle Swarm Optimization, dan K-means, C45 classifier.

BAB III Metodologi dan Desain Sistem

Pada bab ini berisi rancangan dari sistem yang dibangun dan menjelaskan langkah-langkah dari proses awal data kemudian dinormlisasi dan sistem yang dibangun menggunakan dua algoritma yaitu MPSO, Kmeans dan C45 Classifier sebagai algoritma klasifikasi.

BAB IV Pengujian dan Analisis

Bab ini berisi analisis hasil pengujian dari sistem prediksi yang dibangun dengan metode MPSO dan Kmeans.

BAB V Kesimpulan dan Saran

Bab ini berisi kesimpulan dari seluruh tugas akhir dan juga saran yang diperlakukan untuk perkembangan lebih lanjut.