

# 1. PENDAHULUAN

## 1.1 Latar Belakang

Data mining merupakan suatu proses penggalian informasi yang diperoleh dari database. Pada data mining terdapat beberapa tantangan yang dihadapi, salah satunya adalah permasalahan data dimensi tinggi. Data dimensi tinggi merupakan suatu data yang terdiri dari ratusan bahkan ribuan dimensi. Meningkatnya jumlah dimensi data akan mengakibatkan kurangnya performansi pada proses data mining. Permasalahan data dimensi tinggi ini disebut "*Curse of Dimensionality*" [5]. Salah satu contoh data dengan dimensi tinggi adalah data penyakit.

Dalam menyelesaikan permasalahan data dimensi tinggi, diperlukan adanya reduksi dimensi yang berguna untuk mendapatkan keakurasian data yang lebih baik. Reduksi dimensi dilakukan untuk mengurangi atribut-atribut dari suatu data yang tidak diperlukan. Dalam permasalahan ini diperlukan suatu metode yang dapat membantu dalam melakukan reduksi dimensi.

*Principal Component Analysis* (PCA) merupakan salah satu teknik pengurangan dimensi yang bertujuan untuk mengurangi dimensi pada data *unsupervised* [9]. Pada Tugas akhir ini PCA digunakan sebagai teknik *preprocessing* data, dimana hasil *preprocessing* ini akan digunakan pada algoritma *Differential Evolution*.

*DE* (*Differential Evolution*) merupakan salah satu algoritma yang termasuk dalam *EAs* (*Evolutionary Algorithms*) yang berguna sebagai salah satu teknik optimasi efektif yang biasanya digunakan pada permasalahan ilmiah maupun rekayasa. Pada tugas akhir ini algoritma DE digunakan untuk mendapatkan parameter yang akan digunakan pada *classifier* LS-SVM.

*LS-SVM* (*Least Square Support Vector Machine*), yang merupakan suatu tools pengklasifikasian pada suatu sampel tertentu dari *SVM* (*Support Vector Machine*). Kekurangan dari *LS-SVM* adalah sensitif pada perubahan nilai parameter. Algoritma *DE* digunakan untuk pengoptimasian parameter pada *LS-SVM* [1]. Penelitian dengan menggunakan DE dan *LS-SVM* sudah pernah dilakukan sebelumnya oleh Omar S. Skoliman dan Eman AboElHamd dalam sebuah jurnal internasional berjudul "*Classification of Breast Cancer using Differential Evolution and Least Squares Support Vector Machine*" dengan tingkat akurasi sebesar 99,75% [1].

## 1.2 Perumusan Masalah

Berdasarkan latar belakang diatas, maka terdapat beberapa permasalahan yang akan diselesaikan dalam tugas akhir ini, permasalahan tersebut terdiri dari:

1. Bagaimana implementasi Algoritma *Differential Evolution* dalam mengoptimasi *Least Squares Support Vector Machine* (LS-SVM) pada data penyakit berdimensi tinggi?
2. Bagaimana cara kerja Algoritma *Differential Evolution* untuk menemukan parameter LS-SVM pada data penyakit berdimensi tinggi?
3. Bagaimana performansi yang diperoleh *Least Squares Support Vector Machine* (LS-SVM) pada data penyakit berdimensi tinggi ?

Adapun batasan masalah dari tugas akhir ini adalah sebagai berikut:

1. Data yang digunakan merupakan data penyakit leukimia dan colon tumor yang diambil dari *Biomedical Dataset* pada *Kent Ridge*.
2. Tidak membahas PCA secara mendalam.
3. Tidak dilakukan penanganan *outlier* pada dataset yang digunakan.
4. Dataset sudah memiliki label class.
5. Data yang digunakan merupakan data numerik.

### **1.3 Tujuan**

Berdasarkan perumusan masalah diatas tujuan untuk menyelesaikan masalah tersebut adalah:

1. Mengimplementasikan Algoritma Differential Evolution dan *Least Squares Support Vector Machine (LS-SVM)* untuk memprediksi penyakit pada data berdimensi tinggi.
2. Mengetahui cara kerja Algoritma *Differential Evolution* untuk menemukan parameter LS-SVM pada data penyakit berdimensi tinggi.
3. Menganalisis performansi yang diperoleh *Least Squares Support Vector Machine (LS-SVM)* pada data penyakit berdimensi tinggi.

### **1.4 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 akan menjelaskan tugas secara umum.

#### **BAB II     Kajian Pustaka**

Bab ini menjelaskan uraian teori yang berhubungan dengan data dimensi tinggi, PCA, DE dan LS-SVM.

#### **BAB III    Metodologi dan Desain Sistem**

Pada bab ini berisi rancangan dari sistem yang dibangun, dimana pembangunan sistem ini akan menjelaskan langkah-langkah dari proses PCA serta pembangunan algoritma *hybrid* DE dan LS-SVM dalam memprediksi penyakit.

#### **BAB IV     Pengujian dan Analisis**

Bab ini berisi analisis hasil pengujian dari sistem prediksi penyakit yang dibangun dengan menggunakan metode DE dan LS-SVM. Untuk pengujian sistem yang digunakan telah dibuat pada bab sebelumnya.

#### **BAB V      Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari seluruh tugas akhir dan juga saran-saran yang diperlukan untuk pengembangan lebih lanjut.