

ABSTRAK

Data mining merupakan salah satu teknik yang paling sering digunakan dalam mengolah sebuah data untuk mendapatkan sebuah informasi penting yang terdapat dalam data tersebut. Namun apabila dihadapkan oleh sebuah data yang berukuran sangat besar atau berdimensi tinggi, maka hal tersebut mengakibatkan besarnya sumber daya komputasi yang diperlukan untuk memproses data tersebut dengan waktu yang cepat. Selain itu tingkat optimasi dan efisiensi menjadi sebuah permasalahan juga karena banyaknya jumlah atribut yang ada.

Salah satu solusi untuk melakukan efisiensi pemrosesan yang dilakukan terhadap data berdimensi tinggi adalah dengan menerapkan *Evolutionary Algorithm* (EAs) ke dalam metode data mining.

Oleh karena itu dalam tugas akhir ini penulis akan membangun sebuah system prediksi dengan menerapkan salah satu algoritma EAs yaitu *Genetic Algorithm* yang akan digunakan dalam menyelesaikan permasalahan pengklasifikasian data berdimensi tinggi. Data yang digunakan adalah data set penyakit yang didapatkan dari *website Kent Ridge Bio-medical Data Set Repository* yang masing-masing datanya memiliki ribuan atribut.

Dari sistem tersebut didapatkan hasil yang akan menunjukkan seseorang berpotensi terkena penyakit yang sedang diuji atau tidak. Untuk mengetahui kinerja sistem yang telah dibangun dilakukan perhitungan akurasi. Sehingga hasil akhir dari percobaan ini didapatkan bahwa penggunaan *Genetic Algorithm* dalam menyelesaikan sebuah permasalahan klasifikasi pada data berdimensi tinggi dapat dikatakan berhasil. Hal tersebut dikarenakan sistem yang telah dibangun menghasilkan nilai akurasi yang cukup tinggi yakni 89.47% untuk data set penyakit colon tumor dan 81.82% untuk data set penyakit leukemia.

Kata Kunci : Klasifikasi, Asosiasi Klasifikasi, Algoritma Evolusi, Algoritma Genetika, Data berdimensi tinggi.