Abstrak

Dimensionality adalah salah satu tantangan dalam *data mining,* tantangan ini meliputi jumlah atribut yang begitu besar sehingga sering disebut dengan *curse of dimensionality.* Semakin besar jumlah atribut maka semakin memakan waktu dan memerlukan upaya komputasi yang berlebihan sehingga data sulit untuk ditangani. Hal yang diperlukan untuk mengatasi tantangan ini adalah dengan cara mereduksi dimensi dari data tersebut.

Teknik reduksi yang dibahas pada tugas akhir ini adalah dengan menggunakan algoritma *K-Means* dengan cara pengelompokan data pada setiap *cluster*. Algoritma ini digunakan untuk mereduksi *record* yang kemudian dilanjutkan oleh GA sebagai *feature selection* untuk memilih atribut-atribut yang paling optimal berdasarkan nilai *fitness* tertinggi. Pencarian nilai *fitness* dilakukan dengan menggunakan metode klasifikasi yaitu SVM.

Hasil dari pengujian sistem menghasilkan data yang direduksi oleh *K-Means* memiliki akurasi yang lebih rendah untuk *dataset* tertentu dibandingkan tanpa menggunakan *K-Means.* Atribut optimal yang dihasilkan GA bervariasi berdasarkan parameter yang digunakan. Data yang digunakan adalah data penyakit berdimensi tinggi berupa ekspresi gen yaitu *colon tumor* dan *leukemia.* Akurasi rata-rata terbaik yang didapat pada data *colon tumor* adalah 92.86% dengan jumlah atribut terpilih yaitu 983 atribut, sedangkan untuk data *leukemia* selalu menghasilkan atribut yang berkualitas dengan rata-rata akurasi 100%.

Kata kunci : *dimensionality, data mining, K-Means,* GA, SVM