

Abstrak

Data merupakan sumber penting yang dapat diolah menjadi suatu informasi yang berguna. Untuk menghasilkan informasi yang berguna data harus diolah dengan baik. *Data mining* merupakan proses mengolah data menjadi suatu informasi. Dalam *data mining* terdapat berbagai macam teknik untuk mengolah data, salah satunya adalah *clustering*. *Clustering* merupakan proses pengelompokan data berdasarkan kesamaan karakteristik yang dimiliki data, sehingga dihasilkan beberapa *cluster* dengan karakteristik yang berbeda. Untuk mengukur kesamaan antardata dapat digunakan pengukuran jarak untuk data bertipe numerik, misalnya *Euclidean distance*. *Euclidean distance* sangat mudah digunakan dan berkeja sangat baik ketika data bertipe numerik, tetapi tidak dapat digunakan ketika data bertipe kategori atau campuran. Di dunia nyata banyak data yang bertipe kategori maupun campuran. Pada Tugas Akhir ini akan dilakukan *clustering* data bertipe campuran dengan menggunakan *Incremental Genetic K-means Algorithm* (IGKA).

Pada Tugas Akhir ini perhitungan jarak akan dilakukan secara terpisah untuk data numerik dan kategori. Algoritma *Incremental Genetic K-means* dimulai dengan inialisasi individu sejumlah populasi P secara random, lalu dilakukan perhitungan fitness dan seleksi, kemudian individu yang terseleksi akan dimutasi. Setelah tahap mutasi selesai, maka semua individu masuk ke tahap k-means. Proses ini akan terus berjalan sampai akhir iterasi atau jika pusat sudah tidak berubah. Dari hasil pengujian dan analisa didapatkan bahwa jumlah klaster dapat mempengaruhi *silhouette coefficient*. Semakin besar jumlah klaster maka *silhouette coefficient* akan semakin besar.

Kata kunci: data mining, *Clustering*, IGKA, data bertipe campuran