

# 1. Pendahuluan

## 1.1 Latar belakang

Data Mining merupakan salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari database skala besar sebagai tuntutan dari pertumbuhan teknologi informasi. Dimana data mining itu sendiri adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari kumpulan data [8]. *Clustering* merupakan salah satu fungsionalitas dari data mining yang digunakan dalam mengekstrak pengetahuan yang berguna untuk mendapatkan pola yang menarik dari volume data yang banyak dan dimensi data yang besar. Dimana *clustering* mengklasifikasikan data kedalam kelas atau *cluster* yang memiliki persamaan dan perbedaan yang dibawa oleh masing-masing atribut data. Klaster yang baik adalah klaster yang memiliki persamaan (*similarity*) intraklaster yang tinggi dan perbedaan (*dissimilarity*) antarklaster yang tinggi.

Saat ini banyak aplikasi yang menggunakan klasterisasi dalam pemecahan masalahnya sehingga banyak pendekatan yang ditawarkan dengan algoritmanya masing-masing, diantaranya adalah metode Partisi (K-means), Hierarki, Fuzzy C-means, dll. *Clustering* dengan metode K-means yang dikembangkan oleh Mac Queen pada tahun 1967, sangat terkenal dengan kemampuannya untuk mengklaster data yang besar dan dapat menangani data outlier. K-means merupakan metode pengklasteran yang memisahkan data kedalam  $k$  kelompok yang berbeda artinya sebelum dilakukan klasterisasi maka user harus menentukan jumlah  $k$  partisi yang diinginkan. Selain itu pendekatan umum dari klasterisasi adalah menemukan titik pusat klaster yang merepresentasikan tiap klaster. Oleh karena itu K-means juga melakukan penentuan titik pusat klaster yang dibangkitkan dengan cara random. Hanya saja dalam penentuan titik pusat tersebut K-means masih sangat sensitif. K-means akan mampu menemukan titik pusat yang tepat apabila pembangkitan awal titik pusat yang dilakukan dengan random tersebut mendekati solusi akhir pusat klaster begitu juga sebaliknya. Jika awal titik pusat jauh dari solusi akhir pusat klaster maka kemungkinan besar hasil klasterisasinya menjadi tidak tepat. Dari keterangan tersebut diketahui bahwa K-means hanya dapat mencapai local optimal saja, belum mampu mencapai global optimalnya.

Selain algoritma K-means, beberapa peneliti juga menentukan klaster dengan algoritma genetika (Genetic Algorithm, disingkat GA). Dimana ide dasarnya adalah untuk mensimulasikan proses evolusi dari seleksi alami dan mengembangkan solusi dari satu generasi ke generasi berikutnya. Jika dibandingkan dengan K-means maka akan menjadi sangat kontras karena GA tidak sensitif pada inisialisasi awal dan selalu konvergen pada wilayah global. Hanya saja biaya komputasi menggunakan GA ini mahal untuk aplikasi yang luas. Ketidakefisienan waktu pada GA ini disebabkan karena GA menggunakan crossover operator yang membutuhkan waktu yang lama untuk menghasilkan kromosom anak yang valid dari kromosom induknya. Crossover operation sangat rumit dan membutuhkan perulangan yang tidak sedikit untuk menghasilkan kromosom legal. Selain itu juga membutuhkan biaya mahal pada perhitungan

fungsi fitnessnya. Oleh karena itu Khrisna dan Murty mencoba menggabungkan kekuatan alami GA dan kesederhanaan dari K-means menjadi algoritma *Genetic K-means Algorithm* (GKA) [9]. Dimana operator crossover pada GA digantikan dengan operator K-Means yang mengambil langkah serupa dengan *K-Means Algorithm* sebagai operator pencarian. Dan dari hasil penelitian mereka diperoleh hasil bahwa algoritma GKA lebih cepat dibandingkan menggunakan algoritma genetika murni karena GKA selalu menghasilkan konvergensi pada wilayah global.

Begitu juga dengan FGKA (*Fast Genetic K-means Algorithm*) yang merupakan pengembangan dari algoritma GKA dimana keduanya mampu menghasilkan konvergensi pada wilayah global [11]. Hanya saja performansi algoritma FGKA jauh lebih cepat dibandingkan dengan GKA. Oleh karena itu, hal tersebutlah yang menjadi alasan pemilihan algoritma FGKA dalam memecahkan kasus *clustering* sehingga diharapkan dapat diperoleh hasil klaster yang optimal dan dengan performansi semaksimal mungkin.

## 1.2 Perumusan masalah

Dengan mengacu pada latar belakang masalah di atas, maka permasalahan yang akan dibahas dan diteliti adalah :

1. Bagaimana menerapkan algoritma *FGKA (Fast Genetic K-means Algorithm)* pada metode *Clustering*.
2. Bagaimana performansi Algoritma *FGKA (Fast Genetic K-means Algorithm)* pada *Clustering* sebagai suatu metode dalam menentukan *cluster* dengan efektif dan menghasilkan *cluster* yang lebih optimal dibandingkan dengan menggunakan metode yang sejenis dan standar lainnya.

## 1.3 Tujuan

Berdasarkan pada masalah yang telah diidentifikasi di atas, maka tujuan Tugas Akhir ini adalah:

1. Mengimplementasikan metode *Clustering* dengan *FGKA (Fast Genetic K-means Algorithm)*.
2. Menganalisis hasil klasterisasi yang dihasilkan oleh perangkat lunak *Clustering* menggunakan *FGKA (Fast Genetic K-means Algorithm)* dan membandingkannya dengan hasil klasterisasi menggunakan metode K-means, yang mana parameter pembandingnya adalah akurasi (rata-rata error) dan *time performance* .

## 1.4 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Studi Literatur  
Mencari referensi dan sumber-sumber lain yang layak yang berhubungan dengan *data mining*, *Clustering*, *Genetic Algorithm* dan *FGKA (Fast Genetic K-means Algorithm)*.
2. Pendalaman Materi

Mempelajari konsep clustering dan algoritma *Fast Genetic K-means Algorithm* sehingga dapat menentukan tujuan yang ingin dicapai berdasarkan parameter-parameter inputan.

3. Perancangan dan Implementasi  
Merancang program dengan perancangan terstruktur dan mengimplementasikan hasil perancangan menggunakan bahasa pemrograman Delphi 7.
4. Analisis dan Evaluasi  
Melakukan pengujian perangkat lunak dengan menganalisa performansi metode Klasterisasi dengan *Fast Genetic K-means Algorithm* berdasarkan parameter input berupa jumlah klaster yang diinginkan ( $k$ ), ukuran populasi ( $M$ ), jumlah maksimum generasi ( $G$ ), dan besarnya probabilitas mutasi yang digunakan dalam penentuan klaster dari masing-masing data.
5. Penyusunan Laporan Tugas Akhir  
Menyusun laporan hasil analisa yang dirangkum ke dalam sebuah buku Laporan Tugas Akhir.