

1. Pendahuluan

1.1 Latar Belakang

Teknologi komputasi dan media penyimpanan telah memungkinkan manusia untuk mengumpulkan dan menyimpan data dari berbagai sumber dengan jangkauan yang amat luas. Meskipun teknologi basis data modern telah menghasilkan media penyimpanan yang ekonomis bagi aliran data yang amat besar ini, teknologi untuk membantu kita menganalisis, memahami, atau bahkan memvisualisasikannya belum tersedia. Hal inilah yang melatar belakangi dikembangkannya konsep pengambilan pengetahuan dari basis data atau data mining yang dapat diterjemahkan secara sederhana sebagai 'penambangan data'[4].

Di dalam data mining, data yang mempunyai sifat dan karakteristik yang berbeda dari data pada umumnya dan mempunyai kemunculan kejadian relatif sedikit dikatakan sebagai *outlier*[8]. *Outlier detection* dapat didefinisikan sebagai pencarian terhadap sebagian kecil dari data, yang memiliki sifat yang berbeda jika dibandingkan dengan data keseluruhan. *Outlier* sendiri dapat didefinisikan sebagai sebuah titik data pada suatu basis data yang sangat berbeda dibandingkan dengan titik data pada basis data pada umumnya[9].

Permasalahan yang ada adalah metode-metode yang ada sebelumnya tidak mengkhususkan pencarian *outlier* pada data yang bersifat *categorical* (tipe data yang tidak dapat dilakukan perhitungan aritmetik). Adapun metode yang dapat menyelesaikan masalah diatas yaitu *LSA(Local Search Algorithm)*, namun metode tersebut masih dirasa kurang optimal dalam hal performansi waktunya, karena banyak sekali dalam kehidupan nyata kecepatan pencarian data outlier sangat dibutuhkan. Maka salah satu metode yang dapat digunakan untuk memperbaiki permasalahan tersebut yaitu metode Algoritma *Fast Greedy*. Karena algoritma ini lebih optimal dalam perhitungan waktu eksekusi dibandingkan dengan algoritma yang sudah ada (*Local Search Algorithm*), tanpa mengabaikan tingkat akurasi.

1.2 Perumusan Masalah

Permasalahan yang dihadapi pada Tugas Akhir ini sebagai berikut :

1. Bagaimana implementasi *Fast Greedy Algorithm* pada aplikasi *outlier detection*.
2. Bagaimana analisis akurasi dan analisis waktu *Fast Greedy Algorithm* dalam menentukan *outlier*.
3. Bagaimana membandingkan tingkat performansi waktu dan akurasi penggunaan *Fast Greedy Algorithm* dengan *LSA(Local Search Algorithm)*.

Adapun batasan masalah dalam Tugas Akhir ini adalah :

1. Algoritma yang digunakan adalah *Fast Greedy Algorithm*.
2. Tidak menangani *preprocessing* data.
3. Tipe data berupa *categorical*.
4. Input data yang akan dilakukan deteksi *outlier* adalah file *.arff*.
5. Pengujian dilakukan pada dataset yang diketahui jumlah *outliernya*.
6. Analisis performansi meliputi akurasi obyek *outlier* yang dihasilkan, dan waktu eksekusi terhadap peningkatan jumlah data dan penambahan jumlah inputan *k-outlier*.

1.3 Tujuan

Tujuan dari pembuatan Tugas Akhir ini adalah :

1. Implementasi *Fast Greedy Algorithm* untuk mendeteksi *outlier* pada data *categorical*.
2. Melakukan analisa terhadap perangkat lunak untuk menguji akurasi dan waktu deteksi terhadap penambahan jumlah data dan nilai inputan *k-outlier*.
3. Mendapatkan perbandingan tingkat performansi penggunaan *Fast Greedy Algorithm* dengan *LSA(Local Search Algorithm)*.

1.4 Metodologi Penyelesaian Masalah

Metodologi yang akan digunakan dalam merealisasikan tujuan dan pemecahan masalah di atas adalah dengan menggunakan langkah-langkah berikut :

1. Studi pustaka
Pada tahap ini dipelajari metoda *Fast Greedy Algorithm* dalam dektesi *outlier* pada *categorical* data, dan *LSA (Local Search Algorithm)*.
2. Analisis dan Desain
Pada tahap ini dilakukan analisis pemecahan dari permasalahan pada data *categorical* yang akan dideteksi *outlier*-nya dengan menggunakan dasar teori yang telah dipelajari pada tahap sebelumnya.
3. Implementasi
Hasil yang telah dilakukan pada tahap perancangan dapat diimplementasikan pada bahasa pemrograman.
4. Pengujian dan Evaluasi
Menganalisis dengan mengukur akurasi dan waktu pendeteksian *outlier* hasil implementasi metoda *Fast Greedy Algorithm*.