

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Perkembangan yang pesat dalam teknologi pengumpulan dan penyimpanan data membuat data yang cukup besar dapat terakumulasi dengan mudah. Mengambil informasi yang berguna dari data tersebut pun menjadi hal yang sangat menantang. Teknik atau alat tradisional dalam mengolah data juga tidak dapat digunakan karena besarnya ukuran dari kumpulan data. Banyak data tersimpan dalam berbagai bidang seperti bisnis, obat-obatan, dan ilmu pengetahuan.

Data mining diperlukan untuk melakukan pengolahan data dengan jumlah yang banyak dan besar. Data mining adalah teknologi yang menyatukan teknik tradisional dalam menganalisis data dengan algoritma yang canggih untuk memproses data (Tan, Steinbach and Kumar 2006). Data mentah yang jumlahnya banyak dan besar tidak hanya didiamkan, tetapi melalui proses data mining data tersebut dapat menghasilkan informasi yang berguna. Data mining membuka kesempatan mengetahui sesuatu yang baru dari data yang lama, dengan cara yang baru.

Metode *graph-based clustering* adalah salah satu metode data mining yang fleksibel. Metode ini dapat digunakan untuk data yang sifatnya tidak teratur atau bahkan data yang teratur, untuk data yang memiliki kesamaan atau tidak memiliki kesamaan antara satu dengan yang lainnya (Tan, Steinbach and Kumar 2006).

Dataset yang memiliki hubungan antar *instance* salah satunya adalah data jadwal. Data jadwal memiliki hubungan dari segi waktu dan tempat. Data jadwal dapat difokuskan lagi kedalam berbagai macam data salah satunya adalah data penjadwalan kuliah. Data penjadwalan kuliah adalah data jadwal setiap mata kuliah yang ada pada periode tertentu. Data tersebut menyediakan hari dan jam dimana mata kuliah akan dijalankan, dosen yang mengajar, tempat, dan informasi lain seperti jumlah sks setiap mata kuliah, semester mata kuliah, dan jenis mata kuliah. *Graph-based clustering* yang dapat dilakukan terhadap data penjadwalan

kuliah. Hasil *clustering* dapat melihat pengelompokan yang terjadi terhadap jadwal dan juga ruangan. Hasil tersebut dapat digunakan untuk pengambilan keputusan di masa depan dalam pembuatan jadwal kuliah.

### 1.2. Rumusan Masalah

Berdasarkan latar belakang yang terjadi maka dapat diambil dua rumusan masalah dalam penelitian ini. Rumusan masalah tersebut adalah:

1. Bagaimana melakukan analisis data mining menggunakan *graph-based clustering* dengan algoritma *simple k-means*.
2. Bagaimana implementasi perangkat lunak untuk melakukan data mining dengan *graph-based clustering* dengan algoritma *simple k-means*.

### 1.3. Tujuan

Berdasarkan rumusan masalah yang ada maka dapat diambil dua tujuan dari penelitian ini. Tujuan tersebut adalah:

1. Melakukan analisis *graph-based clustering* terhadap kelompok data menggunakan algoritma *simple k-means* dengan perubahan parameter *seed* dan jumlah *cluster*.
2. Implementasi perangkat lunak untuk melakukan data mining dengan *graph-based clustering* menggunakan algoritma *simple k-means*.

### 1.4. Batasan Masalah

Penelitian data mining menggunakan metode *graph-based clustering* ini perlu dibatasi agar tidak meluas. Batasan dari penelitian ini diantaranya adalah:

1. Aplikasi memproses file data dengan format *.arff* dan *.csv*.
2. Aplikasi melakukan proses data mining dengan *graph-based clustering* tanpa memperhitungkan waktu pemrosesan.
3. Aplikasi menggunakan *library Weka*.
4. Aplikasi melakukan analisis *cluster* terhadap data jadwal perkuliahan Fakultas Teknologi Informasi Semester Genap 2013/2014 dan Semester Ganjil 2014/2015 sebagai data studi kasus.

### 1.5. Sistematika Pembahasan

Laporan penelitian mengenai *graph-based clustering* ini dibagi menjadi ke dalam enam bab besar. Setiap bab terbagi kedalam beberapa sub bab. Sistematika pembagian bab tersebut adalah:

Bab I adalah bab pendahuluan. Bab ini membahas latar belakang pembuatan penelitian data mining menggunakan *graph-based clustering*, masalah-masalah yang diangkat, tujuan dari penelitian, batasan dari penelitian, dan sistematika laporan penelitian.

Bab II adalah bab landasan teori. Bab ini memberikan informasi teori yang digunakan dalam penelitian. Teori analisis maupun implementasi. Bab ini memberikan informasi teori analisis mengenai data mining, apa itu data mining, data, *clustering*, *graph-based clustering*, dan *Weka*.

Bab III adalah bab analisis dan disain. Bab ini menjabarkan analisis mengenai *graph-based clustering* dan disain perancangan aplikasi. Bab ini terdiri dari analisis proses *graph-based clustering*, gambaran dari antarmuka aplikasi, antarmuka perangkat keras, antarmuka perangkat lunak, disain penyimpanan data, dan disain antarmuka aplikasi.

Bab IV adalah bab pengembangan perangkat lunak. Bab ini menjabarkan implementasi dari perangkat lunak yang dibuat dalam penelitian. Bab ini terdiri dari implementasi class, implementasi penyimpanan data, dan implementasi antarmuka aplikasi.

Bab V adalah bab testing dan evaluasi sistem. Bab ini melakukan testing terhadap aplikasi yang dibuat. Bab ini terdiri dari rencana pengujian yang akan dilakukan terhadap aplikasi dan pelaksanaan pengujian dengan metode *white box* dan *black box testing*.

Bab VI adalah bab kesimpulan dan saran. Bab terakhir ini akan membahas kesimpulan yang dihasilkan dari penelitian ini dan saran untuk pengembangan analisis data mining dengan *graph-based clustering* dan pengembangan aplikasi.