

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi yang semakin besar telah mempengaruhi aktivitas pengelolaan data dan informasi di berbagai bidang yang mengharuskan adanya penyimpanan informasi di dalamnya. Informasi-informasi yang dulu disimpan dengan melakukan pencatatan dalam format kertas (*paper based*) kini semakin banyak dikembangkan ke basis elektronik dengan dukungan teknologi informasi dan sistem komputerisasi.

Semakin besarnya jumlah informasi yang tersedia dalam bentuk elektronik, seperti publikasi elektronik, *e-mail*, dan berbagai jenis dokumen elektronik menyebabkan perkembangan ukuran *database* teks menjadi sangat cepat. Saat ini sebagian besar informasi yang ada dalam suatu pemerintahan, industri, atau lembaga lainnya disimpan dalam *database* dokumen, yang terdiri dari kumpulan besar dokumen dari berbagai jenis, seperti artikel, makalah, buku, catatan, pesan dan lain-lain (Han, J dan Kamber, M., 2006 : 614). Dari banyaknya data pada *database* dokumen maka semakin besar pula ruang penyimpanan dari dokumen-dokumen tersebut dan semakin sulit pula dalam mencari informasi yang ada pada suatu dokumen. Tanpa mengetahui informasi apa saja yang terdapat pada setiap dokumen-dokumen yang ada akan sulit untuk mencari informasi yang diinginkan (Han, J dan Kamber, M., 2006 : 615).

Data yang disimpan dengan tidak terklasifikasi dan kemungkinan pemakaian ruang penyimpanan data oleh beberapa orang yang berbeda menyebabkan sulitnya mencari informasi-informasi yang berkaitan. Hal tersebut akan lebih mudah dilakukan apabila pada *database* dokumen telah dilakukan pengelompokan-pengelompokan dokumen sesuai dengan kesamaan-kesamaan yang dimiliki. Dari pengelompokan-pengelompokan tersebut dapat ditemukan perbedaan-perbedaan dari setiap kelompok, besarnya keperluan dan relevansi dari suatu dokumen, atau penemuan pola-pola yang ada pada dokumen-dokumen.

Pengelompokan dokumen yang jumlahnya besar akan sulit dilaksanakan apabila dilakukan secara manual. Oleh karena itu diperlukan metode untuk melakukannya secara otomatis. *Document clustering* merupakan suatu metode yang dapat digunakan untuk mengelompokkan dokumen secara otomatis. Tujuannya yaitu untuk menemukan pola pengelompokan dari data-data dokumen. Dokumen-dokumen yang memiliki kesamaan akan berada pada satu *cluster*, sedangkan dokumen-dokumen yang berbeda akan berada pada *cluster* yang berbeda (Rahmawati, 2007).

Metode yang banyak digunakan untuk masalah *document clustering* ini yaitu algoritma *K-Means*. Terdapat banyak varian dari algoritma *K-Means* yang banyak diantaranya dapat mengatasi kelemahan dari algoritma *K-Means* yang biasa. Salah satu dari varian algoritma *K-Means* yaitu *Bisecting K-Mean*. Metode *Bisecting K-Means* menghasilkan pengelompokan yang lebih baik secara signifikan dan cukup konsisten sesuai dengan kesamaan *entropy* serta kualitas *cluster* secara keseluruhan (Steinbach, M., Karypis, G., Kumar, V., 2000 : 18).

## 1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah:

1. Bagaimana membangun suatu sistem yang dapat melakukan pengelompokan berkas-berkas dokumen elektronik?
2. Bagaimana metode *Bisecting K-Means* dapat membantu proses pengelompokan dokumen?
3. Bagaimana hasil pengelompokan yang diperoleh dari metode *Bisecting K-Means*?

## 1.3 Tujuan

Tujuan yang ingin dicapai dalam skripsi ini yaitu:

1. Membuat suatu sistem yang dapat mengelompokkan berkas-berkas dokumen elektronik.
2. Menggunakan metode *Bisecting K-Means* untuk membantu proses pengelompokan dokumen elektronik.
3. Mengetahui kualitas hasil pengelompokan dokumen yang dihasilkan oleh metode *Bisecting K-Means*.

## 1.4 Batasan Masalah

Dalam pembuatan skripsi ini, diterapkan beberapa batasan masalah yaitu:

1. Data yang di olah untuk pengelompokan dokumen adalah data teks dari setiap dokumen elektronik yang bertipe file *DOC*, *DOCX*, *PPT*, *PPTX*, serta *PDF* .
2. Data teks pada dokumen elektronik yang digunakan merupakan teks berbahasa Indonesia.

## 1.5 Metode Penelitian

Metodologi yang digunakan dalam penelitian ini adalah:

1. Eksplorasi dan Studi Literatur

Dalam mempelajari konsep-konsep pembangunan sistem pengelompokan dokumen, peneliti melakukan eksplorasi dengan cara membaca serta membandingkan literatur-literatur seperti jurnal, karya ilmiah, *textbook*, *paper* dan sumber ilmiah lainnya.

2. Analisis dan Perancangan Perangkat Lunak

Analisis dan perancangan dilakukan dengan menentukan bahasa pemrograman yang akan digunakan, struktur data, input/output serta algoritma yang akan digunakan.

3. Implementasi dan Pengujian

Implementasi dilakukan setelah analisis dilakukan. Pengujian kualitas dilakukan dengan membandingkan hasil keluaran dengan kenyataan pada sample.

4. Hasil akhir dan Penarikan Kesimpulan

Analisis hasil akhir dilakukan untuk mengelompokan dokumen elektronik menggunakan metode *Bisecting K-Means*. Jika hasilnya kurang memuaskan, maka dilakukan analisis akhir untuk mengetahui penyebabnya, selanjutnya dilakukan penarikan kesimpulan.

## 1.6 Sistematika Penulisan

Sistematika dalam penyusunan skripsi ini adalah sebagai berikut :

### Bab I Pendahuluan

Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metode penelitian serta sistematika penulisan yang digunakan dalam penyusunan skripsi.

### Bab II Tinjauan Pustaka

Berisi dasar teori yang digunakan dalam analisis, perancangan, dan implementasi perangkat lunak.

### Bab III Metodologi Penelitian

Berisi tentang teknis pelaksanaan penelitian berupa alat dan bahan penelitian, desain penelitian dan proses penelitian.

### Bab IV Hasil Penelitian dan Pembahasan

Berisi hasil penelitian serta analisis yang dilakukan selama penelitian.

### Bab V Kesimpulan dan Saran

Berisi kesimpulan yang didapat selama penelitian dan saran-saran dalam meningkatkan kualitas dan kuantitas hasil penelitian.

### Lampiran

Berisi dokumen-dokumen yang menunjang keabsahan penelitian.