

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pertumbuhan digital tekstual dalam mendistribusikan informasi pada saat ini, berkembang dengan cepat. Terutama pertukaran informasi berbasis *web*. Banyak lembaga media yang memanfaatkannya sebagai sarana publikasi. Salah satunya adalah lembaga penyedia berita. Para penerbit surat kabar saat ini banyak yang menyajikan beritanya dalam versi *online*. Hal ini tentunya memudahkan seseorang untuk memperoleh berita dengan cepat dan mudah. (Karhendana, 2008).

Ramainya lembaga berita yang menyediakan artikel dalam bentuk *online* memang memudahkan pengguna dalam mengetahui peristiwa terkini. Namun, melimpahnya *content* yang dihasilkan oleh setiap situs berita dapat menyulitkan pengguna dalam mencerna semua informasi yang dihasilkan. Akibatnya, pengguna dapat terjebak dalam arus berita yang melimpah (Karhendana, 2008).

Agregator berita adalah sistem yang dikembangkan untuk mengatasi permasalahan tersebut. Agregator adalah sebuah sistem atau perangkat lunak, umumnya berupa aplikasi *web*, yang “mengagregasi” (menggabungkan) berbagai *content* elektronik dari berbagai sumber, seperti artikel berita, blog, maupun *podcast*. Agregator dapat mengurangi waktu dan usaha yang dibutuhkan untuk mengakses *update* terbaru dari sebuah situs. Agregator berita (*news aggregator*) merupakan salah satu jenis agregator yang mengumpulkan artikel berita dari

berbagai sumber, kemudian menyajikannya kembali kepada pengguna dalam satu kesatuan (Karhendana, 2008). Dengan demikian, pengguna tidak perlu lagi menjelajah ke berbagai situs berita hanya untuk mengetahui peristiwa yang baru terjadi.

Akan tetapi, sebuah agregator berita “polos” memiliki kelemahan, yaitu menggabungkan semua berita hasil perolehannya dalam satu kumpulan besar. Hal ini dapat menyebabkan efek *information overload* bagi pengguna, karena belum tentu pengguna tertarik dengan semua berita yang terdapat pada agregator berita. Seringkali pengguna hanya tertarik untuk mengetahui peristiwa apa saja yang baru terjadi (Karhendana, 2008).

Masalah tersebut dapat diatasi dengan mengelompokkan berita sesuai peristiwa yang diulasnya. Pengelompokkan berita yang dilakukan secara manual mahal, oleh karena itu diperlukan metode untuk mengelompokkan data secara otomatis. *Document clustering* adalah salah satu metode yang dapat digunakan untuk mengelompokkan dokumen secara otomatis. Tujuan dari proses *document clustering* ini adalah menemukan pola pengelompokan yang alami dari data. Dimana dokumen teks yang memiliki persamaan akan berada dalam satu *cluster* dan dokumen yang memiliki banyak perbedaan akan berada dalam *cluster* yang berbeda (Rahmawati, 2007).

Terdapat banyak algoritma *clustering* yang telah dikembangkan untuk berbagai keperluan. Tentunya setiap algoritma tersebut memiliki kelebihan dan kekurangan masing-masing. Dalam hal ini pengelompokkan dokumen berita menggunakan algoritma *K-Means*. Algoritma *K-Means* merupakan algoritma

dasar yang mudah dan dapat dengan cepat melakukan pemusatan data (*convergence*) pada proses *clustering* (Adiningsih, 2007).

Saat ini, sudah banyak agregator berita yang telah memanfaatkan *document clustering*. Akan tetapi masih sedikit sistem yang mengikutsertakan berita berbahasa Indonesia. Oleh karena itu, dalam skripsi ini penulis merancang sebuah aplikasi yang menerapkan metode *document clustering* untuk mengelompokkan berita berbahasa Indonesia sebagai solusi dari permasalahan yang telah dipaparkan di atas, skripsi ini diberi judul “Penggunaan Algoritma K-Means untuk *Clustering* Dokumen Berita Berbahasa Indonesia”.

1.2 Rumusan Masalah

Bagaimana algoritma K-Means diterapkan sehingga dapat mengelompokkan dokumen berita berbahasa Indonesia?

1.3 Batasan Masalah

Dalam pengerjaan skripsi ini, ditetapkan beberapa batasan masalah, antara lain sebagai berikut:

1. Perangkat lunak yang dikembangkan adalah berbentuk *prototype*.
2. Artikel yang digunakan sebagai sumber adalah dokumen berita yang menggunakan bahasa Indonesia.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam pelaksanaan skripsi ini adalah mengimplementasikan algoritma *K-Means* yang dapat digunakan untuk mengelompokkan dokumen berita.

1.5 Manfaat Penelitian

1. Aplikasi ini dapat menjadi solusi untuk memecahkan permasalahan pengelompokan secara otomatis.
2. Memberikan informasi kepada khalayak tentang penggunaan algoritma K-Means untuk mengelompokkan dokumen berita berbahasa Indonesia.

1.6 Metode Penelitian

Metode yang diterapkan dalam pembuatan skripsi ini, antara lain :

1. Eksplorasi dan Studi Literatur

Eksplorasi dan studi literatur dilakukan dengan mempelajari konsep-konsep yang berkaitan dengan skripsi ini, seperti *clustering* pada aliran berita yang terdiri dari karakteristik berita dan transformasi berita menjadi dataset, algoritma *Clustering*, melalui literatur-literatur seperti buku (*textbook*), paper, dan sumber ilmiah lain seperti situs internet ataupun artikel dokumen teks yang berhubungan.

2. Analisis dan Perancangan Perangkat Lunak
3. Implementasi Program dan Pengujian Performansi

Analisis dan perancangan perangkat lunak dilakukan untuk menentukan permasalahan mengenai bahasa pemrograman yang akan digunakan, struktur data, *input/output* program, dan permasalahan teknik algoritma yang akan diimplementasikan.

4. Eksperimen terhadap algoritma

Eksperimen terhadap algoritma dilakukan untuk mengetahui bagaimana algoritma bekerja dengan berbagai perlakuan.

5. Perbaikan

Perbaikan dilakukan terhadap kesalahan-kesalahan yang mungkin terjadi pada perangkat lunak, laporan, maupun dokumentasi teknis.

6. Hasil Akhir dan Penarikan Kesimpulan

Analisis hasil dilakukan untuk mengetahui performansi pembangunan perangkat lunak, jika ternyata hasilnya baik, maka dilakukan analisis akhir untuk mengetahui penyebabnya, selanjutnya dilakukan penarikan kesimpulan.

1.7 Sistematika Penelitian

Dalam penyusunan skripsi ini, sistematika penulisan dibagi menjadi beberapa bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas masalah yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian ini memuat landasan teori yang berfungsi sebagai sumber atau alat dalam memahami permasalahan yang berkaitan dengan teori dan konsep *Data Mining*, *Clustering*, *Text Mining* dan *Document Clustering*, algoritma *K-Means*, serta Agregator Berita.

BAB III METODOLOGI PENELITIAN

Bab ini merupakan penjabaran dari alat dan bahan penelitian, metode pengembangan sistem yang digunakan.

BAB IV HASIL PENELITIAN

Bab ini akan membahas secara mendalam hal-hal rumusan masalah untuk mencari solusi atas rumusan masalah tersebut.

BAB V PENUTUP

Bab ini akan diutarakan kesimpulan yang merupakan jawaban atas rumusan masalah dalam penelitian, serta memberikan saran yang dapat dijadikan rekomendasi dalam melakukan penelitian selanjutnya.

