

BAB I

PENDAHULUAN

1.1. Latar Belakang

Semakin meningkatnya penggunaan internet di era digital ini memberikan dampak positif. Salah satunya adalah memudahkan dalam bertukar maupun mencari informasi. Dengan bertambahnya jumlah karya ilmiah yang bervariasi setiap tahunnya, menyebabkan terjadi penumpukan dokumen. Rata-rata *user* tidak ingin membaca isi dari keseluruhan karya ilmiah karena membutuhkan waktu yang lama. Oleh karena itu, dibutuhkan suatu aplikasi yang dapat memberikan kemudahan dalam pencarian data tertentu.

Klasterisasi merupakan salah satu metode *data mining* yang bersifat *unsupervised learning* yang mengelompokkan data ke dalam suatu kelas, sehingga objek pada cluster memiliki kesamaan dengan objek lain pada *cluster* sama [1]. Klasterisasi telah diterapkan dalam beberapa bidang seperti *biometric*, *web mining*, *information retrieval* dan lain-lain. Beberapa metode *clustering* telah diterapkan dalam pengelompokkan teks terstruktur sejenis *relational database* seperti *k-means*, *decision tree*, *Self Optimization Maps* (SOM) dan sebagainya. Data berjumlah besar akan menyebabkan besarnya matrik term-dokumen dan proses clustering tidak optimal. Hal tersebut karena adanya data yang tidak relevan. Oleh karena itu, dibutuhkan suatu metode yang dapat menangani permasalahan clustering dokumen.

K-Means adalah metode *clustering* (pengelompokan) data yang terkenal simple dan cepat dalam perhitungannya [2]. Alasan menggunakan metode *k-means* adalah karena metode ini mempunyai kualitas cukup tinggi terhadap ukuran objek, sehingga metode ini relatif lebih efisien dalam mengolah objek dalam jumlah besar.

Dari latar belakang di atas, maka inilah alasan penulis mengambil tema *cluster* dokumen teks abstraksi untuk memberikan pengetahuan tentang pengelompokkan dokumen sesuai dengan *cluster* masing-masing.

1.2. Perumusan Masalah

Berdasarkan penjelasan dari latar belakang, peneliti dapat mengidentifikasi perumusan masalah yaitu seberapa tingkat akurasi metode *k-means* dalam *clustering* klasterisasi karya ilmiah abstrak.

1.3. Tujuan

Berdasarkan perumusan masalah tersebut, maka tujuan penelitian yang ingin dicapai adalah menganalisa tingkat akurasi hasil klasterisasi karya ilmiah dengan menggunakan metode *k-means*.

1.4. Manfaat

Beberapa manfaat yang dapat diperoleh dari hasil penelitian ini adalah sebagai berikut:

1. Memberikan gambaran tentang algoritma *k-means* pada penerapan dokumen teks.
2. Memberikan kontribusi bagi pengembangan ilmu pengetahuan dan teknologi serta menambah wawasan pengetahuan bagi pengembangan dan penelitian selanjutnya khususnya dalam bidang *Data Mining*.

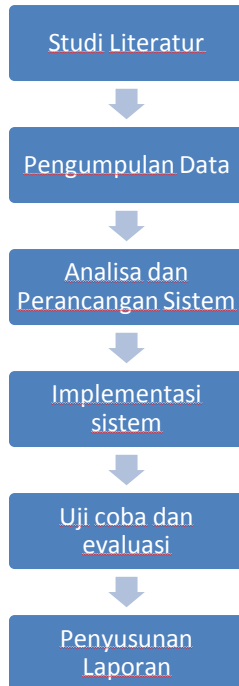
1.5. Batasan Masalah

Agar penelitian ini tercapai dengan baik, maka dibuat suatu batasan masalah sebagai berikut:

1. Data yang digunakan dalam uji coba implementasi pada *clustering* adalah karya ilmiah abstrak skripsi Universitas Trunojoyo Madura.
2. Dokumen yang digunakan untuk proses *clustering* telah mengalami *preprocessing*.
3. Perhitungan *distance* (jarak) pada metode *k-means* menggunakan *Euclidean distance*.

1.6. Metodologi Penelitian

Untuk menyelesaikan pembuatan skripsi, akan dilakukan beberapa langkah metodologi penelitian. Berikut adalah langkah-langkah metodologi penelitian pada skripsi ini.



Gambar 1.1. Rancangan Penelitian

1. Studi literatur

Studi literatur dilakukan dengan mempelajari teori yang berkaitan dengan topik Skripsi seperti *clustering*, *information retrieval* dan *text mining* melalui literatur buku, jurnal, serta sumber ilmiah lainnya melalui situs *internet* ataupun artikel yang berkaitan.

2. Pengumpulan data
Penelitian ini menggunakan data primer abstraksi skripsi. Data yang dikumpulkan pada penelitian digunakan sebagai bahan dalam melakukan pengujian *cluster*. Karena data yang diperoleh akan dijadikan acuan dalam mengambil kesimpulan.
3. Analisis dan Perancangan sistem
Pada tahap ini dilakukan analisis terhadap permasalahan yang terjadi serta analisa terhadap kebutuhan dari sistem. Sedangkan perancangan bagaimana merancang hasil analisis mendesain sebuah sistem untuk dapat diimplementasikan.
4. Implementasi
Pada tahap ini dilakukan proses implementasi pengkodean program dalam aplikasi komputer dengan menggunakan bahasa pemrograman yang telah dipilih sesuai dengan analisis dan perancangan yang sudah dilakukan.
5. Uji coba dan evaluasi
Dari *output* program yang telah dibuat, maka perlu adanya pengujian terhadap sistem yang dibuat dalam implementasi. Pengujian ini dilakukan dengan mengetahui apakah system tersebut berjalan sesuai dengan analisis dan perancangan. Kemudian dari pengujian tersebut akan evaluasi kembali berdasarkan hasil *clustering*.
6. Penyusunan buku skripsi
Tahap terakhir ini mengambil kesimpulan dari hasil penelitian dan dibukukan pada laporan Tugas Akhir.

1.7. Sistematika Penulisan

BAB I : Pendahuluan

Bab ini berisikan pendahuluan yang meliputi latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB II : Tinjauan Pustaka

Bab ini membuat pembahasan mengenai penelitian terdahulu yang digunakan sebagai referensi dalam penulisan penelitian ini. Selain itu juga memuat teori-teori yang berhubungan dengan pembahasan Skripsi seperti *text mining* dan *k-means*.

BAB III : Perancangan Sistem

Bab ini berisi tentang analisis dan perancangan sistem yang diperlukan dalam menganalisis permasalahan yang ada serta merancang dari analisis pengolahan data dengan menggunakan metode yang ditentukan.

BAB IV : Implementasi Sistem

Bab ini dijelaskan mengenai implementasi pembuatan aplikasi berdasarkan hasil rancangan yang telah dibuat sebelumnya. Aplikasi yang telah selesai nantinya akan diuji dan dievaluasi untuk perbaikan sistem.

BAB V : Penutup

Bab ini berisikan tentang kesimpulan dari hasil penelitian yang telah dilakukan serta menjelaskan saran-saran penulis kepada pembaca agar aplikasi yang dirancang dapat dikembangkan lagi.

[Halaman Ini Sengaja Dikosongkan]