

1. Pendahuluan

1.1 Latar belakang

Saat ini jumlah dokumen teks berbentuk digital semakin banyak. Dokumen-dokumen tersebut menawarkan berbagai informasi yang beragam. Untuk memudahkan pengambilan informasi yang dibutuhkan dari dokumen teks tersebut, maka perlu dilakukan pengelompokan dokumen sesuai dengan topiknya. Pengelompokan ini dapat dilakukan dengan menggunakan teknik yang terdapat dalam *data mining* yaitu *clustering*. *Clustering* adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota dalam satu kelompok dan meminimumkan kesamaan antar kelompok [16]. *Clustering* merupakan teknik yang bersifat *unsupervised learning*, yaitu data tidak diberi label sehingga data menemukan pola distribusinya sendiri. Dengan demikian, algoritma *clustering* akan bekerja untuk mengelompokkan dokumen berdasarkan kesamaan anggotanya [9].

Terdapat tantangan yang harus dihadapi dalam menyelesaikan masalah pengelompokan dokumen, antara lain jumlah dokumen yang besar dengan dimensi data yang tinggi sehingga algoritma *clustering* yang digunakan harus mampu menghasilkan kompleksitas waktu dan ruang yang efisien. Selain itu, algoritma yang digunakan pun harus dapat menghasilkan *cluster* dengan kualitas yang baik. Oleh karena itu, dalam tugas akhir ini akan dibangun sistem pengelompokan dokumen dengan menggunakan algoritma *committee clustering*. Algoritma ini akan membangun pusat *cluster* dengan merata-ratakan nilai *feature vector* dari himpunan bagian dari anggota *cluster* yang disebut *committee*. *Committee* ini, nantinya akan bertugas untuk menentukan suatu dokumen untuk masuk ke dalam suatu *cluster*. Dengan memilih anggota *committee* secara hati-hati, *feature* dari pusat *cluster* akan cenderung mengarah kepada *class* target [8]. Dengan demikian, algoritma ini akan mampu menghasilkan kualitas *cluster* yang baik.

1.2 Perumusan masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diambil rumusan masalah sebagai berikut :

1. Bagaimana melakukan pengelompokan dokumen dengan menggunakan algoritma *committee clustering*?
2. Bagaimana mengetahui dan menganalisis kualitas *cluster* yang dihasilkan dengan menggunakan metode *silhouette coefficient* dan menganalisis parameter-parameter dari algoritma *committee clustering* yang berpengaruh terhadap kualitas *cluster*?

1.3 Batasan Masalah

Adapun batasan-batasan masalah dalam pengerjaan Tugas Akhir ini adalah:

1. Tugas akhir ini menggunakan dokumen berita berbahasa Indonesia yang telah mengalami *proses text preprocessing* terlebih dahulu.
2. Kualitas *cluster* yang dihasilkan tidak dibandingkan dengan algoritma lain.

1.4 Tujuan

1. Melakukan pengelompokkan dokumen dengan menggunakan algoritma *committee clustering*.
2. Mengetahui dan menganalisis kualitas *cluster* yang dihasilkan dengan menggunakan metode *silhouette coefficient* dan menganalisis parameter-parameter dari algoritma *committee clustering* yang berpengaruh terhadap kualitas *cluster*.

1.5 Metodologi penyelesaian masalah

Metodologi yang digunakan untuk menyelesaikan permasalahan-permasalahan dalam Tugas Akhir ini terdiri dari 6 tahap, yaitu:

1. Studi pustaka
Pada tahapan ini akan dilakukan pendalaman materi mengenai algoritma *committee clustering*.
2. Tahap pengumpulan data dan *preprocessing* data
Pada tahapan ini dilakukan pencarian dokumen berita berbahasa Indonesia. Setelah itu, dilakukan proses *preprocessing* terhadap dokumen sehingga dokumen tersebut dapat digunakan sebagai masukan untuk sistem yang akan dibangun.
3. Tahap perancangan sistem
Pada tahapan ini dilakukan proses perancangan sistem dengan membuat gambaran fungsi-fungsi yang akan membangun sistem pengelompokkan dokumen dengan algoritma *committee clustering*.
4. Tahap implementasi pemrograman
Pada tahapan ini dilakukan proses pembangunan sistem berdasarkan rancangan yang telah dibuat sebelumnya dengan menggunakan *tools matlab*.
5. Tahap pengujian sistem
Pada tahapan ini dilakukan proses pengujian sistem yang menggunakan sejumlah dokumen dengan mengganti-ganti nilai parameter k dan nilai *threshold*, dan dievaluasi dengan menggunakan metode *silhouette coefficient*.
6. Tahap analisis output sistem
Pada tahapan ini dilakukan analisa kualitas *cluster* yang dihasilkan oleh sistem.
7. Tahap Pembuatan Laporan.
Pada tahap ini dilakukan penyusunan laporan akhir dan pengumpulan dokumentasi berdasarkan analisa hasil pengujian tugas akhir ini.

1.6 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika sebagai berikut:

1. Pendahuluan

Bab ini menguraikan tugas akhir ini secara umum, meliputi latar belakang, perumusan masalah, batasan masalah, tujuan dan metodologi penyelesaian masalah.

2. Dasar Teori

Bab ini membahas mengenai uraian teori yang berhubungan dengan *clustering* dan algoritma *committee clustering*.

3. Analisis Perancangan Dan Implementasi

Bab ini berisi analisis kebutuhan dari sistem yang kemudian dituangkan ke dalam suatu sistem pemodelan secara terstruktur. Dari tahap analisis kemudian dilanjutkan ke tahap perancangan dan implementasi.

4. Analisis Hasil Pengujian

Bab ini membahas mengenai pengujian yang dilakukan terhadap sistem yang telah dibangun. Pengujian dilakukan dengan mengganti-ganti nilai parameter yang terdapat dalam sistem. Tahap pengujian dilanjutkan dengan tahap analisis hasil pengujian.

5. Kesimpulan

Berisi kesimpulan dari penulisan Tugas Akhir ini dan saran-saran yang diperlukan untuk pengembangan lebih lanjut.