

# BAB 1 PENDAHULUAN

## 1.1. Latar Belakang

Sistem pengelompokan dokumen berita merupakan sistem yang menggabungkan dokumen berita berdasarkan tingkat kemiripannya. Sistem ini akan menyusun dokumen berita secara terstruktur dan rapi sesuai dengan kelompoknya. Kelompok-kelompok berita yang terbentuk akan memberikan kemudahan untuk membaca berita. Pembaca berita akan dimudahkan dalam memilih dan mencari berita-berita yang diinginkan.

Sistem pengelompokan ini akan dikelompokkan menggunakan metode *single linkage Clustering* dan *k-means Clustering*. Metode *single linkage* merupakan teknik pengelompokan yang bekerja berdasarkan prinsip dari Algoritma Hierarchical *Clustering*. Sedangkan *K-means* merupakan teknik pengelompokan yang bekerja berdasarkan *Partitioned Clustering*. Prinsip kerja dari pengelompokan Hierarchical *Clustering* dilakukan secara bertahap. Dan disetiap iterasi dari pengelompokan hierarchial *Clustering* hanya ada satu pemilihan penggabungan suatu dokumen terhadap dokumen lainnya. Sedangkan prinsip kerja dari pengelompokan *Partitioned Clustering* adalah mengelompokkan dokumen secara acak karena dipengaruhi centroid. Dan disetiap iterasi dari pengelompokan *Partitioned Clustering* memungkinkan untuk terjadinya lebih dari satu pemilihan dokumen-dokumen yang akan digabungkan.

Sistem pengelompokan dokumen ini akan membandingkan performansi dari dua metode tersebut. Parameter yang digunakan untuk membandingkan performansi metode tersebut adalah *silhouette coefficient* dan *purity*. *Silhouette coefficient* diperoleh dengan menghitung jarak rata-rata antar dokumen dalam satu *cluster*. Setelah itu menghitung jarak antar suatu dokumen dengan dokumen lain yang berada dalam *cluster* lain dan yang diambil adalah jarak yang paling terdekat. Sedangkan pengujian *purity*, akan dilakukan secara manual. Pengujian manual yang dimaksud adalah, dengan memeriksa setiap dokumen dalam suatu *cluster* apakah merupakan dokumen yang mirip atau berada dalam satu kategori yang sama.

## **1.2. Rumusan Masalah**

Rumusan masalah dalam tugas akhir ini adalah :

1. Bagaimana membuat suatu sistem yang bisa mengelompokkan dokumen menggunakan metode *Single linkage* dan *K-means*.
2. Bagaimana melakukan pengujian untuk melihat metode yang lebih baik antara *Single linkage* dan *K-means* berdasarkan *Silhouette coefficient* dan *Purity*.

## **1.3. Tujuan**

Tujuan dari tugas akhir ini adalah :

1. Mengimplementasikan metode *Single linkage* dan *K-means*.
2. Membandingkan metode antara *Single Linkage* dan *K-means* untuk mengetahui metode yang lebih baik dalam mengelompokkan dokumen

## **1.4. Batasan Masalah**

Batasan masalah yang digunakan dalam tugas akhir ini adalah sebagai berikut :

1. Dataset yang digunakan adalah dokumen berita yang diambil dari internet.
2. Aplikasi bekerja secara offline
3. Stemming yang digunakan adalah stemming bahasa Indonesia sehingga tidak mengatasi kata asing yang terdapat dalam dokumen
4. Dokumen yang digunakan hanya dalam bentuk format .txt.

## 1.5. Metodologi Penelitian

Metode yang digunakan dalam menyelesaikan tugas akhir ini adalah menggunakan metode studi pustaka atau studi literature dan analisis dengan langkah kerja sebagai berikut :

1. Mencari dan mempelajari referensi bahan – bahan yang berhubungan dengan tugas akhir ini seperti Text Mining, Hierarchical Agglomerative *Clustering*, Partitioned *Clustering*, Single linkage, *K-means*.
2. Merancang aplikasi untuk melakukan pengelompokan data dan mengimplementasikannya kedalam perangkat lunak
3. Melakukan pengujian sistem dengan data yang diperoleh
4. Melakukan analisis dari hasil pengujian
5. Membuat kesimpulan dari hasil implementasi dan analisis
6. Menyusun laporan tugas akhir

## 1.6. Sistematika Penulisan

Tugas akhir ini disusun berdasarkan sistematika sebagai berikut

### **BAB I : Pendahuluan**

Pada bab ini berisi latar belakang masalah, perumusan masalah yang akan dibahas, batasan masalah, tujuan yang akan dicapai, metodologi penyelesaian, serta sistematika penulisan.

### **BAB II : Dasar Teori**

Pada bab ini berisi dasar teori yang digunakan dalam membangun sistem untuk Tugas akhir ini.

### **BAB III : Analisis Dan Perancangan Sistem**

Pada bab ini berisi analisis sistem yang meliputi gambaran umum dan analisis kebutuhan sistem, serta perancangan sistem

### **BAB IV : Implementasi Dan Pengujian**

Pada bab ini akan diuraikan mengenai hasil yang didapatkan dari *Clustering* dokumen otomatis menggunakan metode Single Linkage dan *K-means* dan akan dilakukan analisis parameter evaluasi hasil *Clustering* dan nilai *purity*-nya.

### **BAB V : Penutup**

Bab ini akan berisi kesimpulan dan saran dari hasil pengujian yang dilakukan serta diberikan saran-saran untuk pengembangan lebih lanjut perangkat lunak ini.