

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pemanfaatan teknologi digital telah menjadi kebutuhan dalam era modern saat ini.

Dengan semakin berkembangnya teknologi informasi menyebabkan semakin mudahnya pertukaran data digital, dalam mempermudah menemukan, dan menganalisis informasi – informasi penting yang berjumlah ratusan bahkan ribuan, sehingga dalam praktek kenyataannya banyak orang yang kesulitan bahkan mengeluh dalam kasus seperti ini.

Analisa klaster merupakan alat analisis *statistic multivariate* dan merupakan salah satu metode penting dalam pengenalan pola tidak terbimbing (*unsupervised pattern recognition*), dalam pemaparan analisis metode klaster ini sangat luas karena metode dari *unsupervised data* pola yang tidak terbimbing ini sangat erat hubungannya dengan konteks teks mining.

Dalam bidang sistem informasi temukembali analisa klaster dengan konteks teks mining, metode clustering juga telah diterapkan pada berbagai kasus, misalnya dalam mengekstrak konsep (Karypis, 2000), mempartisi corpus dokumenteks (Grossman and Frieder, 2004), pengkategorian tingkat similarity dokumen dengan penggunaan kata kunci (Herwansyah, 2009). Dalam teknik pengklasteran ditemukan kesulitan untuk mengklaster dokumenteks, dimana dokumenteks memiliki banyak kata – kata yang saling berkaitan antara makna satu dengan makna yang lainnya, sehingga asumsi dalam mengelompokkan dokumenteks jika dalam dokumen kata – kata yang muncul saling relevan antar dokumen satu dengan dokumen lainnya, maka jarak antar dokumen akan *similar* atau menghasilkan kemiripan yang tinggi, sehingga klaster dapat terbentuk akurat, tetapi jika kata yang muncul tidak relevan antar dokumennya, maka jarak yang dihasilkan pun tidak

similar, sehingga pengklasteran memiliki tingkat kesulitan yang tinggi, sehingga klaster yang dihasilkan kurang akurat.

Oleh karena itu sistem akan dibuat dan di rancang untuk membuktikan apakah dengan menggunakan similarity *Jaccard Coefficient* dan *similarity* menggunakan jarak *Euclidian Distance* akan menghasilkan pengklasteran yang lebih baik,. Diharapkan dari hasil penelitian ini didapatkan clustering dokumen yang memiliki tingkat kesamaan yang tinggi dan tingkat relevansi klaster yang lebih baik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, penulis dapat merumuskan masalah yang akan dibahas yaitu, bagaimana merancang bangun aplikasi dokumen clustering dengan metode k-means berdasarkan winnowing fingerprint similarity.

## 1.3 Batasan Masalah

Agar penekanan tujuan dalam penelitian ini tepat mencapai sasaran, maka penulis membatasi masalah yang akan dibahas sebagai berikut:

1. Dokumen teks uji berupa dokumen teks latin dengan format plaintext (txt dan dtxt).
2. Penentuan input Cluster dokumen dilakukan oleh inputan dari pengguna.
3. Perhitungan Similaritas menggunakan *Jaccard Coefficient* dan *Euclidian Distance*.

## 1.4 Tujuan

Tujuan dari penelitian ini dapat diuraikan sebagai berikut:

1. Menerapkan metode winnowing untuk mendapatkan fingerprint dokumen.
2. Menerapkan metode clustering k-means yang lebih baik untuk hasil pengelompokan dengan menggunakan *Jaccard coefficient* dan *Euclidian distance*.

3. Menganalisis, merancang, dan mengimplementasikan aplikasi dokumen clustering berdasarkan tingkat similaritinya, sehingga didapatkan hasil klaster sesuai dengan anggota dokumen tekstual lainnya.

## **1.5 Sistematika Penulisan**

Sistematika penulisan tugas akhir ini dibagi menjadi 6 (enam) bab. Setiap bab terdiri dari sub-sub bab dan penjelasan yang tersusun sehingga mudah untuk dipahami. Berikut penjelasan tentang masing-masing bab:

### **Bab I Pendahuluan**

Merupakan deskripsi umum dari tugas akhir ini, yang meliputi: latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika pembahasan tugas akhir.

### **Bab II Landasan Teori**

Pada bab ini berisikan mengenai teori-teori yang berhubungan dengan pembahasan tugas akhir ini, seperti: *Text Mining*, *Clustering*, *Algoritma winnowing*, *Algoritma K-Means*.

### **Bab III Metodologi Penelitian**

Pada bab ini berisikan pedoman tentang tahapan-tahapan yang dilakukan dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan yang telah ditentukan sebelumnya.

### **Bab IV Analisis Dan Perancangan**

Pada bab ini berisikan tentang pembahasan mengenai analisis perangkat lunak yang terdiri dari: analisis dan perancangan *Aplikasi*.

### **Bab V Implementasi Dan Pengujian**

Pada bab ini berisikan penjelasan mengenai implementasi perangkat lunak yang meliputi: alasan pemilihan perangkat lunak, batasan implementasi, lingkungan implementasi dan implementasi antar muka serta menjelaskan pengujian program ini.

### **Bab VI Kesimpulan Dan Saran**

Pada bab ini berisikan kesimpulan dari tugas akhir yang dibuat dan menjelaskan saran-saran penulis kepada pembaca agar Aplikasi yang dibuat dapat dikembangkan lagi.