

# 1. Pendahuluan

## 1.1 Latar Belakang Masalah

Dunia pendidikan tidak terlepas dari tugas dan ujian. Dengan banyaknya tugas yang diberikan, terkadang membuat pengajar sulit untuk memeriksa apakah tugas yang dikerjakan oleh pelajar merupakan hasil sendiri atau meniru hasil tugas pelajar yang lain. Banyak cara yang digunakan pelajar untuk meniru hasil tugas pelajar yang lain, seperti melakukan perubahan susunan kalimat, mengganti suatu kata dengan sinonimnya, merubah kalimat aktif menjadi pasif, atau hanya menambah sedikit kata-kata. Hal inilah yang mempersulit pengajar dalam mengetahui apakah pelajar tersebut mengerjakan tugasnya secara mandiri atau tidak. *Clustering* dapat menjadi solusi dari masalah tersebut, dimana *clustering* dapat digunakan untuk menganalisis data dengan cara mengelompokkan objek ke dalam kelompok-kelompok berdasarkan suatu kemiripan tertentu sehingga semua anggota dari setiap partisi mempunyai kemiripan[5].

Sebuah *cluster* adalah sekumpulan objek yang digabung bersama karena kemiripan atau kedekatannya. Teknik *Clustering* sangat berguna karena akan mentranslasi ukuran persamaan yang intuitif menjadi ukuran yang kuantitatif. Dalam tugas akhir ini, metode *clustering* yang digunakan adalah CLHM (*Centroid Linkage Hierarchical Method*). Metode ini akan mengelompokkan dokumen berdasarkan nilai *centroid* terdekat. Untuk jumlah *cluster*-nya *user* tidak perlu menentukan berapa jumlah yang tepat, karena dengan metode *Hill Climbing*, jumlah cluster akan dipilih sehingga dihasilkan jumlah *cluster* secara otomatis.

Metode *Hill Climbing* ini berfungsi untuk melakukan identifikasi terhadap pergerakan varians yang dihasilkan dari tiap tahap pembentukan *cluster*. *Hill Climbing* akan mencari pada tahap mana terdapat *global optimum*, dengan cara menganalisis pola varians tersebut. Pengelompokkan dengan CLHM dan proses *Hill Climbing Automatic Clustering* sangat memudahkan *user* karena menghasilkan *cluster* secara otomatis dan tepat.

## 1.2 Perumusan Masalah

Berdasarkan pada latar belakang masalah diatas, maka permasalahan yang diangkat dalam tugas akhir ini ialah jumlah dokumen yang tersedia sekarang sangatlah banyak, dan memerlukan waktu yang sangat lama apabila ingin mengelompokkannya secara manual sehingga diperlukan sebuah sistem yang dapat mengelompokkan dokumen tersebut.

Batasan masalah pada tugas akhir ini adalah:

- a. Dataset yang digunakan adalah dokumen abstrak tugas akhir dengan jumlah 50 dokumen.
- b. Aplikasi bekerja secara *offline*
- c. *Stemming* yang digunakan adalah stemming bahasa Indonesia sehingga tidak mengatasi kata asing yang terdapat dalam dokumen
- d. Dokumen yang digunakan hanya dalam bentuk format .txt.

### **1.3 Tujuan**

Secara umum tujuan dari yang ingin dicapai dalam tugas akhir ini adalah sebagai berikut:

1. Mengetahui pengaruh *Hill Climbing* dalam *clustering* dokumen bahasa Indonesia.
2. Melakukan analisis hasil *clustering* metode *Hill Climbing*.

### **1.4 Metodologi Penyelesaian Masalah**

Metode yang digunakan dalam menyelesaikan tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur 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*, *centroid linkage*, *Hill Climbing*, dan bahan lain yang berhubungan dengan tugas akhir ini
2. Melakukan pencarian data yang akan dikelompokkan
3. Merancang aplikasi untuk melakukan pengelompokan data dan mengimplementasikannya ke dalam perangkat lunak
4. Melakukan pengujian sistem dengan data yang diperoleh
5. Melakukan analisis dari hasil pengujian
6. Membuat kesimpulan dari hasil implementasi dan analisis
7. Menyusun laporan tugas akhir

### **1.5 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 *Hill Climbing* 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.