

# 1. PENDAHULUAN

## 1.1 Latar Belakang Masalah

Dalam era yang berkembang saat ini, jumlah informasi yang ada dan tersebar sangatlah banyak. Namun berlimpahnya informasi ini justru menjadi sulit karena umumnya informasi yang ada beraneka macam dan tidak terstruktur sehingga mempersulit para pencari informasi untuk menemukan informasi sesuai dengan yang mereka cari. Hal inilah yang kemudian mendorong meningkatnya kebutuhan untuk menemukan dan mengelola informasi tersebut dengan baik sehingga dapat dihasilkan pengetahuan yang berguna.

Salah satu komponen yang penting dalam pengelolaan informasi untuk pengelompokan berita adalah dengan *clustering*. *Clustering* ada dua macam, yaitu *clustering* tipe hierarki (*hierarchical clustering*) dan *clustering* tipe non hierarki (*partitioned clustering*). Pada tugas akhir ini akan digunakan *hierarchical clustering*. Ide pada *hierarchical clustering* adalah dengan adanya pengelompokan data dengan membuat suatu hirarki, dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan<sup>[7]</sup>.

Salah satu metode dalam *hierarchical clustering* adalah *Hierarchical Agglomerative Clustering*. Pada *Hierarchical Agglomerative Clustering*, setiap obyek atau observasi dianggap sebagai sebuah cluster tersendiri. Dalam tahap selanjutnya, dua cluster yang mempunyai kemiripan digabungkan menjadi sebuah cluster baru demikian seterusnya.

Untuk peningkatan ke-efektifitasan dari *Hierarchical Agglomerative Clustering* pada klasterisasi berita berbahasa Indonesia akan digabungkan dengan pengimplementasian fungsi evaluasi *fuzzy*. Hasil penelitian menunjukkan bahwa *Fuzzy Hierarchical Clustering* mempunyai efisiensi *clustering* yang lebih tinggi dan presisi<sup>[10]</sup>. *Fuzzy hierarchical clustering* mempunyai kemampuan dalam menemukan daerah-daerah berkonsentrasi tinggi dengan metode *Hierarchical Agglomerative Clustering* yang dapat menganalisa dan menggabungkan daerah data dengan cepat. Kemudian, menggunakan fungsi evaluasi untuk menemukan skema *clustering* optimal<sup>[10]</sup>.

## 1.2 Perumusan Masalah

Pada tugas akhir ini, masalah yang akan diselesaikan adalah :

1. Bagaimana penerapan fungsi evaluasi *fuzzy* pada *Hierarchical Agglomerative Clustering* ?
2. Bagaimana proses pembentukan kluster dengan metode *Fuzzy Hierarchical Agglomerative Clustering* dalam proses klasterisasi serta pelabelan *cluster* dari dokumen berita berbahasa Indonesia berhirarki?
3. Bagaimana kualitas hasil klasterisasi dari pengimplementasian metode *Fuzzy Hierarchical Agglomerative Clustering* dalam proses klasterisasi serta pelabelan *cluster* dari dokumen berita berbahasa Indonesia berhirarki?

Batasan masalah pada tugas akhir ini, yaitu :

1. Dataset yang akan digunakan merupakan artikel berita berbahasa Indonesia yang bersifat *offline*.
2. Tidak menangani *preprocessing*

## 1.3 Tujuan

Tujuan dari penelitian tugas akhir ini adalah :

1. Menerapkan fungsi evaluasi *fuzzy* pada metode *Hierarchical Agglomerative Clustering*.
2. Mengetahui serta menerapkan metode *Fuzzy Hierarchical Clustering* dalam proses klasterisasi serta pelabelan *cluster* dari dokumen berita berbahasa Indonesia berhirarki.
3. Mengetahui kualitas hasil klasterisasi dari pengimplementasian metode *Fuzzy Hierarchical Agglomerative Clustering* dalam proses klasterisasi serta pelabelan *cluster* dari dokumen berita berbahasa Indonesia berhirarki.

## 1.4 Metodologi Penyelesaian Masalah

Metodologi yang digunakan untuk penyelesaian masalah pada tugas akhir ini adalah :

### 1. Studi Literatur

Pencarian yang layak mengenai text mining, *Hierarchical Agglomerative Clustering*, dan *Fuzzy Hierarchical Agglomerative Clustering* serta pelabelan *cluster*.

### 2. Desain Penelitian

Untuk mencapai tujuannya, metode *Hierarchical Agglomerative Clustering* dengan *fuzzy* akan diimplementasikan secara langsung oleh peneliti. Dari hasil desain penelitian ini juga nantinya dapat diketahui kinerja *Fuzzy Hierarchical Agglomerative Clustering* ini pada proses klasterisasi berita berbahasa Indonesia serta pelabelan *cluster*.

### 3. Pengujian dan Analisa Hasil

Melakukan proses pengujian terhadap kualitas hasil *Hierarchical Agglomerative Clustering* dengan *fuzzy*, serta melakukan analisa terhadap kelebihan dan keterbatasan *Hierarchical Agglomerative Clustering* dengan *fuzzy* untuk proses klasterisasi berita berbahasa Indonesia serta pelabelan *cluster*.