

BAB V

PENUTUP

5.1 Kesimpulan

Dari analisis yang dilakukan dan hasil yang didapatkan, dapat diambil beberapa kesimpulan, yaitu:

1. Analisis *cluster* hirarkis dapat diterapkan secara dinamis pada platform *web* termasuk dalam hal visualisasi data.
2. Secara umum, langkah-langkah dalam analisis *cluster* Berbahasa Indonesia adalah: *preparing data* → *preprocessing data* → *feature extraction* → *compute distance* → *hirarchial elustering*
3. Analisis *Cluster* hirarkis dokumen Berbahasa Indonesia dapat dilakukan menggunakan library yang sudah ada pada bahasa pemrograman python.
4. Analisis *Cluster* hirarkis dapat diaplikasikan pada dokumen teks Berbahasa Indonesia dengan menggunakan teknik *preprocessing* Bahasa Indonesia, yaitu :
 - a. Menghilangkan *Stopwords* Bahasa Indonesia sesuai dengan list yang ada dalam penelitian Tara.
 - b. Melakukan *Stemming* menggunakan Algoritma *Porter Stemmer* Bahasa Indonesia.

5. Perbandingan waktu pemrosesan antara model yang melalui tahap stemming dan tidak adalah sebesar 0,003 detik untuk *non-stemming* dan 5,518 detik untuk *stemming* atau tepat 5,515 detik.
6. Proses *stemming* tidak terlalu berpengaruh terhadap model yang dihasilkan dengan pertimbangan selisih nilai koefisien korelasi *cophenetic* yang melalui proses stemming dan tidak hanya terpatut kurang dari 0,05.
7. Metode complete linkage cenderung mempunyai koefisien paling tinggi dengan nilai koefisien 0,78 untuk non-stemming dan 0,79 untuk proses yang melalui tahap stemming. Sedangkan ward linkage cenderung memiliki koefisien paling rendah dengan nilai 0,4 untuk non-stemming dan 0,43 untuk proses yang melalui tahap stemming. Sedangkan single linkage sedikit di bawah metode complete linkage dengan nilai 0,6 untuk non-stemming dan 0,62 untuk proses yang melalui tahap stemming.

5.2 Saran

Pada penelitian ini tentu masih banyak kekurangan dan perlu pengembangan yang lebih lanjut. Penelitian ini menyimpan banyak potensi untuk dikembangkan di beberapa area seperti *big data*, *natural language processing*, dan sebagainya. Maka sangat diharapkan bahwa penelitian ini dapat diperbaiki, dikembangkan, dan dikombinasikan oleh penelitian selanjutnya. Untuk

menempurnakan penelitian ini, Berikut adalah beberapa saran yang diajukan oleh peneliti:

1. Hasil dari penelitian ini belum diuji dari sisi skalabilitasnya, sehingga belum dapat dipastikan penelitian ini dapat menangani data dengan jumlah yang besar. Sehingga penelitian selanjutnya diharapkan lebih fokus pada sisi infrastruktur
2. Penelitian ini hanya menggunakan metode *tf-idf* untuk menstrukturisasi dokumen/teks pendek. Ada beberapa opsi lain seperti metode *diffussion wavelet* yang mungkin dapat memberikan hasil yang lebih baik.
3. Beberapa fitur pada *preprocessing* belum diterapkan dalam penelitian ini seperti *part of speech tagging*.
4. Masih menggunakan *relational database*. Peneliti menyarankan menggunakan *non-relational database* karena skalabilitas dan lebih efektif diterapkan pada data analisis.