

BAB VII

KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan hasil penelitian, didapatkan beberapa kesimpulan yang dapat diambil sebagai berikut.

1. Algoritma *Self-Organizing Map* telah dapat diimplementasikan dalam basis data terdistribusi HBase untuk mengklaster dua *dataset* abstrak penelitian.
2. Pada proses *map-reduce*, impor data membutuhkan lebih sedikit baik waktu maupun memori yang dibutuhkan daripada penghitungan frekuensi kata karena pada proses tersebut hanya dilakukan proses *map* sedangkan *reduce* dijalankan di dalam HBase.
3. Dari parameter yang diujikan dalam penelitian ini untuk *dataset* dengan perbandingan ukuran 1,85 : 180,77 dan perbandingan jumlah *record* 1014 : 84354 menghasilkan perbandingan waktu antara dua *dataset* dengan rata-rata 71,98 : 5474,72 sedangkan untuk parameter simpangan baku sebesar 73,12 : 8556,51.
4. Hasil analisis waktu kompleksitas, algoritma yang diterapkan memiliki notasi asimptotik sebesar $T(k, r) = O(k^2r) + O(k^2) + O(r)$ dengan k sebagai kata kunci dan r sebagai jumlah *record*.
5. Setelah dilakukan pengujian korelasi, parameter waktu komputasi dan besaran simpangan baku antarklaster memiliki keterhubungan yang tidak terlalu tinggi sebesar -0,52 untuk *dataset* internet dan -0,63 untuk *dataset* ETD.
6. Berdasarkan grafik pada bab sebelumnya, semakin banyak jumlah kata kunci yang dimasukkan serta jumlah klaster yang ingin dibentuk maka

semakin lama pula waktu komputasinya dan semakin kecil besaran simpangan baku antarklaster.

7.2 Saran

Beberapa saran yang dapat diberikan penulis apabila penelitian dengan ranahan yang sama akan dilanjutkan diantaranya sebagai berikut.

1. Pada penelitian ini ketiga proses utama masih berjalan sendiri-sendiri, alangkah baiknya pada penelitian selanjutnya proses-proses tersebut dapat dijalankan dalam satu perintah.
2. Penelitian ini menggunakan *dataset* yang sejenis, untuk selanjutnya bisa dipakai untuk data yang memiliki ekstensi berbeda-beda seperti teks, gambar, video dan lain-lain.
3. Parameter yang diuji dapat diperbanyak seperti jumlah perulangan, tingkat pembelajaran, jumlah kata kunci yang di masukkan, variasi jumlah klaster yang dibentuk dan lain-lain.