

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil percobaan yang telah dilakukan, dapat diketahui bahwa metode *Vector Space Model* dan *Clustering* untuk deteksi kesamaan dokumen berbahasa Indonesia memiliki akurasi sangat kecil yaitu *recall* 77,32%, *precision* 65,68% dan *f-measure* 68,31%. dibandingkan dengan penelitian Miguel Sanchez-perez yaitu *recall* 89,57%, *precision* 91,25%, *PlagDet* 90%. Hal ini dikarenakan untuk setiap dokumen yang akan dibandingkan dibutuhkan nilai parameter berbeda agar menghasilkan akurasi yang lebih baik dan kurangnya data set yang diujikan. Lalu waktu eksekusi terbilang cukup lama yaitu rata-rata sebesar 16,65 detik untuk dokumen yang diujikan dengan kata kisaran 150-300 kata.

5.2 Saran

Berdasarkan hasil pengujian yang telah dilakukan, saran yang dianjurkan untuk penelitian selanjutnya yaitu sebagai berikut.

1. Menambahkan lebih banyak data set yang telah diketahui mana yang terindikasi sama atau tidak untuk menghasilkan akurasi yang lebih baik.
2. Menguji dengan beberapa nilai parameter untuk mengetahui nilai parameter yang optimal untuk digunakan dalam pendeteksian kesamaan dokumen dalam bahasa Indonesia.
3. Menambahkan klasifikasi tipe kesamaan berupa *exact copy* atau *obfuscation* (rangkuman).

