

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang dilakukan didapatkan kesimpulan sebagai berikut:

- 1 Sistem deteksi plagiasi menggunakan algoritme Frequency Based Hashing-S pada *file* PDF berhasil dibangun. Sistem dibangun sesuai dengan perancangan dan berjalan dengan baik sesuai dengan skenario dan kebutuhan yang sudah ditetapkan.
- 2 Berdasarkan pengujian kinerja sistem, program yang dibuat sudah efisien dengan rata-rata waktu eksekusi sebanyak 0,36 detik. Untuk proses unggah dataset dan proses unggah datatest, didapat kesimpulan bahwa semakin banyak jumlah *chunk* maka waktu eksekusi semakin lama begitu juga sebaliknya. Proses unggah dataset lebih lama dengan rata-rata waktu eksekusi 4,5 detik dibandingkan dengan proses unggah datatest yang memiliki rata-rata waktu eksekusi 3.0 detik. Hal ini dikarenakan pada proses unggah dataset terdapat dua perintah yaitu menggunggah *file* ke tabel

kosin_proses_dataset dan mengunggah *chunk* ke tabel kosin_proses_tf. Sementara pada proses unggah datatest hanya terdapat satu perintah, yaitu mengunggah *file* ke tabel kosin_proses_datatest. Untuk jumlah *file* yang berada di dataset, didapatkan kesimpulan bahwa semakin banyak jumlah *file* di dataset maka waktu eksekusi akan semakin lama begitu juga sebaliknya.

- 3 Permasalahan tentang penyaringan *file* PDF berdasarkan tingkat kesamaan/plagiasi dapat teratasi dengan mengimplementasikan algoritme Frequency Based Hashing-S. Pengimplementasian algoritme Frequency Based Hashing-S pada *file* PDF telah terbukti memiliki akurasi yang tinggi berdasarkan hasil pengujian menggunakan metode *Confusion Matrix*, dengan hasil *F-Score* 92%, *Recall* 100% dan *Precision* 86%
- 4 Permasalahan algoritme Approximate Matching sekarang yang rentan terhadap serangan aktif dapat teratasi dengan pengimplementasian algoritme Frequency Based Hashing-S. Hal ini dapat dibuktikan berdasarkan hasil pengujian integritas data menggunakan metode *Collision Attack*. Sehingga data dapat terjamin integritasnya tanpa diubah

oleh pihak yang tidak bertanggung jawab.

5.2 Saran

Berdasarkan kesimpulan yang sudah dijelaskan, peneliti memberikan saran agar mengimplementasikan *Optical Character Recognition* yang lebih mendetail. Hambatan pada penelitian ini adalah skema pengkodean menggunakan library PyPDF sebagai OCR masih kurang memuaskan untuk mengenali objek yang terdapat pada *file* PDF. Dokumen PDF menggunakan penyandian seperti UTF-8, ASCII, Unicode, dll. Mengekstrak teks yang berada pada PDF dapat mengakibatkan hilangnya Sebagian data karena skema pengkodean tersebut. Oleh sebab itu, dengan menggunakan OCR yang lebih mendetail menggunakan *Deep Learning* seperti penggunaan algoritme *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), dan *Artificial Neural Network* (ANN) diharapkan mampu meningkatkan akurasi perhitungan similarity.