

## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan pengujian fungsionalitas sistem dan pengukuran akurasi yang telah dilakukan, sistem ekstraksi informasi menggunakan algoritma *Maximum Entropy Markov Model (MEMM)* telah berhasil dibangun dengan perolehan akurasi token-kelas sebesar 77%. Berikut penyebab terhadap akurasi yang diperoleh. Perolehan akurasi token-kelas disebabkan oleh penggunaan fitur pembobotan yang tidak signifikan, sehingga algoritma *MEMM* tidak dapat melakukan klasifikasi secara benar.

#### **5.2 Saran**

Pada penelitian ini masih memiliki beberapa kekurangan, sehingga nilai akurasi pada konsep token-kelas maupun kelas-token tidak dapat diperoleh dengan maksimal. Dengan demikian, beberapa saran akan dipaparkan untuk pengembangan lebih lanjut mengenai sistem ekstraksi informasi menggunakan *machine learning*, diantaranya :

1. Menggunakan data masukan dengan format .docx atau .html agar dapat menggunakan beberapa fitur tambahan seperti mendeteksi *bold*, *italic*, *underline*, dan beberapa *font style* lainnya.
2. Diperlukan suatu fitur tambahan yang signifikan untuk pembobotan kategori Fakultas, Program Studi, dan Universitas. Hal tersebut dilakukan agar ketika proses *testing* dilakukan, algoritma *MEMM* dapat menentukan kelas yang tepat untuk Fakultas, Program Studi, atau Universitas.
3. Perlu dibuktikan algoritma *MEMM* dengan perbandingan parameter dalam melakukan ekstraksi informasi pada dokumen karya tulis ilmiah.

Dibutuhkan pengecekan dan pengkoreksian kesalahan ejaan atau typo pada hasil konversi *file PDF*.

