

## **BAB 6**

### **Simpulan dan Saran**

#### **6.1 Simpulan**

Dari hasil pembahasan yang telah dilakukan dan berdasarkan hasil pengujian dapat ditarik beberapa kesimpulan, yaitu:

1. Aplikasi *Email Clustering* dapat membantu pengguna dalam melakukan proses *clustering* pada data email.
2. Akurasi *clustering* dan *labeling* tertinggi adalah 97,08% dengan menggunakan metode K-Means dan pengambilan data *subject* dan *body email*. Metode X-Means *Heuristic* mampu mencapai akurasi sebesar 90,08% dengan pengambilan data *subject* dan *body email*. Metode X-Means *Dynamic* memiliki akurasi 83,83% dengan pengambilan data *subject* dan *body email*. Hasil akurasi dapat dilihat pada sub bab 5.1.10.
3. Pengambilan data *subject* dan *body email* pada proses *clustering* dan *labeling* terbukti dapat membantu meningkatkan akurasi. Rata-rata akurasi yang didapatkan pada setiap proses *clustering* dan *labeling* dengan pengambilan *subject* dan *body email* adalah 90,33%. Hasil akurasi dapat dilihat pada sub bab 5.1.10.
4. Aplikasi *Email Clustering* dapat melakukan proses *clustering* dengan dua fitur pembobotan yaitu *Ranking Term Frequency* yang menghitung bobot dengan cara meranking token-token atau term pada email dan TF-IDF yang menghitung bobot kemunculan term terhadap dokumen email.

#### **6.2 Saran**

Berdasarkan hal-hal yang telah dicapai dalam implementasi, diperoleh beberapa saran, yaitu:

1. Pengembangan algoritma *clustering* dapat dilakukan dengan menambahkan proses *stemming* pada proses *preprocessing*. Untuk menghindari duplikasi token atau term yang memiliki makna sama namun beda penulisan. Namun penggunaan *stemming* ada baiknya hanya dilakukan pada bahasa inggris. Contoh “Send” dengan “Sending” makna nya adalah sama namun berbeda penulisan *term*.

2. Pengembangan algoritma *clustering* dapat dilakukan dengan menambahkan fitur pemilihan opsi, untuk tokenisasi dengan satu kata atau dengan frasa. Untuk melihat perbedaan hasil dari akurasi *clustering*. Contoh frasa “Tugas Akhir” dengan kata “Tugas” memiliki makna yang berbeda. Sehingga dapat mempengaruhi hasil dari *clustering*.
3. Pengembangan algoritma *clustering* dapat dilakukan dengan menambahkan fokus bahasa yang dapat di *clustering*. Bahasa yang dapat ditoleransi oleh aplikasi ini adalah bahasa indonesia dan bahasa inggris.