

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Seleksi fitur menggunakan metode *chi-square* berhasil diimplementasikan pada data untuk menyeleksi fitur yang relevan untuk menyiapkan data yang akan digunakan pembuatan model klasifikasi. Data yang telah dikumpulkan dari hasil *crawling data Twitter* dilakukan praproses data dan *labelling* untuk dihitung nilai kontingensi dari setiap fitur. Dari nilai kontingensi dihitung nilai *chi square* setiap fitur untuk dilakukan penyeleksian fitur
2. Pembobotan *Bag of Words* dan *TF-IDF* berhasil dilakukan pada data yang telah melalui tahap seleksi fitur. Setiap kata dianggap sebagai fitur dan dokumen direpresentasikan sebagai vektor. *Bag of Words* merepresentasikan teks sebagai vektor berdasarkan kemunculan kata-kata dalam dokumen sedangkan *TF-IDF* merepresentasikan teks sebagai vektor berdasarkan bobot kata dalam dokumen dan seluruh korpus teks. Data yang telah dilakukan pembobotan fitur akan dilakukan pembuatan model klasifikasi
3. Perbandingan performa klasifikasi sentimen program MBKM menggunakan algoritma Random Forest dengan pembobotan fitur *Bag of Words* dan *TF-IDF* menggunakan *splitting data 90 : 10* dan *k-fold cross validation* dengan $k=10$ menunjukkan kombinasi pembobotan fitur *TF-IDF* dengan algoritma *Random Forest* merupakan model terbaik. Berdasarkan hasil evaluasi menggunakan kurva *AUC-ROC*, kombinasi algoritma *Random Forest* dengan pembobotan fitur *TF-IDF* merupakan klasifikasi yang sangat baik. Sedangkan kombinasi algoritma *Random Forest* dengan pembobotan fitur *Bag of Words* berdasarkan kurva *AUC-ROC* merupakan klasifikasi yang baik

5.2 Saran

Berdasarkan analisis dan kesimpulan pada penelitian ini, terdapat beberapa saran yang dapat diterapkan untuk penelitian selanjutnya :

1. Meningkatkan kualitas data teks dengan melakukan verifikasi dan koreksi pada tahap *slang word standardization* untuk memaksimalkan perhitungan seleksi fitur menggunakan *chi square* dan pembobotan fitur menggunakan *Bag of Words* dan *TF-IDF* serta memperoleh hasil analisis yang lebih baik.
2. Masukkan kata dari dataset yang memiliki konotasi positif atau negatif yang belum terdapat dalam kamus ke dalam kamus untuk hasil *labelling* yang lebih baik.