

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Setelah melakukan penelitian mengenai implementasi *feature selection* dan pembobotan TFIDF pada klasifikasi *tweets* informasi kehilangan dengan Naïve Bayes, maka penulis mendapatkan beberapa kesimpulan yang selaras dengan tujuan penelitian. Berikut kesimpulan yang didapatkan oleh penulis.

1. Implementasi teknik *feature selection* menggunakan metode Information Gain, Chi-Square, dan Document Frequency untuk mengklasifikasikan informasi kehilangan pada data Twitter berhasil dilakukan dengan cara menghitung nilai fitur untuk setiap metode, mengurutkannya berdasarkan nilai terbesar, menentukan jumlah fitur yang digunakan dengan *threshold*, dan melakukan klasifikasi menggunakan Naive Bayes. Penulis berhasil mengimplementasikan pembobotan fitur menggunakan TFIDF untuk mengklasifikasikan informasi kehilangan pada data Twitter.
2. Implementasi pembobotan fitur menggunakan TFIDF untuk mengklasifikasikan informasi kehilangan pada data Twitter berhasil dilakukan dengan cara membuat matriks TF berdasarkan frekuensi kemunculan kata, menghitung nilai DF dan IDF dari setiap fitur, menghitung bobot TFIDF berdasarkan bobot TF dan nilai IDF, dan melakukan klasifikasi menggunakan Naive Bayes.
3. Perbandingan performa dari klasifikasi informasi kehilangan pada data Twitter menggunakan Naïve Bayes dengan penerapan *feature selection* dan pembobotan TFIDF menunjukkan nilai akurasi tertinggi terjadi pada model kombinasi Chi-Square – TF dan Chi-Square – TFIDF pada jumlah fitur 20% sebesar 85,09%. Sedangkan akurasi paling rendah didapatkan pada kombinasi Document Frequency dan TFIDF dengan jumlah fitur 80% sebesar 78,88%. Penggunaan *feature selection* pada sebagian besar kombinasi model terbukti efektif untuk meningkatkan akurasi klasifikasi Naïve Bayes, meskipun perbedaannya tidak terlalu signifikan. Perbandingan hasil

pengujian dengan pembobotan TF menunjukkan akurasi yang lebih baik pada sebagian besar model kombinasi dibanding dengan pengujian tanpa *feature selection*. Sedangkan pada pengujian kombinasi *feature selection* dengan TFIDF seluruhnya memiliki hasil akurasi yang lebih baik dibanding pengujian tanpa *feature selection*. Penggunaan bobot TFIDF lebih efektif dikombinasikan dengan Chi-Square karena memiliki akurasi yang lebih baik dibanding metode lainnya di semua jumlah fitur yang digunakan. Namun, penggunaan bobot TF secara keseluruhan memiliki akurasi yang lebih baik.

## 5.2 Saran

Dalam pelaksanaan penelitian, penulis menyadari masih banyak kekurangan yang dilakukan dalam penelitian ini. Oleh karena itu, penulis menyampaikan beberapa saran yang dapat dilakukan di kemudian hari. Penelitian selanjutnya dapat menghasilkan sebuah program yang jauh lebih baik dalam hal akurasi. Berikut beberapa saran yang dapat penulis anjurkan.

1. Menambah jumlah dokumen pada setiap kelasnya agar mendapatkan kamus fitur yang lebih banyak sehingga dapat memaksimalkan proses *training* model.
2. Melakukan penanganan terhadap kesalahan penulisan kata pada dataset agar frekuensi kemunculan kata lebih akurat sehingga dapat memaksimalkan penghitungan nilai *feature selection* dan pembobotan TF serta TFIDF.
3. Melakukan penelitian mengenai modifikasi pada penghitungan TFIDF seperti menggunakan normalisasi, *smoothing*, dan penggunaan basis log yang berbeda dengan harapan dapat meningkatkan akurasi.
4. Melakukan penelitian dengan lebih banyak kombinasi pengujian, khususnya pada jumlah fitur yang digunakan agar mendapatkan hasil yang lebih bervariasi sehingga tren pada grafik dapat terlihat lebih jelas.
5. Diharapkan penelitian selanjutnya dapat menghasilkan program dengan antarmuka sehingga dapat digunakan sebagai sistem informasi kehilangan.