

## BAB 5 SIMPULAN DAN SARAN

### 5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan mengenai analisis sentimen dengan algoritma Naive Bayes, dapat ditarik simpulan sebagai berikut.

1. Analisis sentimen mengenai kereta cepat Jakarta-Bandung dan kereta panoramic dengan menggunakan algoritma Naive Bayes berhasil diimplementasikan. Dalam penelitian ini, digunakan CountVectorizer dan TF-IDF untuk pembobotan kata serta SMOTE dan NearMiss untuk penyeimbangan data.
2. Berdasarkan hasil uji coba yang telah dilaksanakan, hasil model pembelajaran mesin yang menghasilkan nilai akurasi, *precision*, *recall*, dan *f1-score* tertinggi didapatkan melalui model yang menggunakan TF-IDF dalam pembobotan kata dengan rasio *train test split* sebesar 70%:30% dan *random state* 42 serta data yang tidak seimbang dan tidak mengalami *oversampling* atau *undersampling*. Selain itu, model pembelajaran mesin dibuat dengan menggunakan algoritma Complement Naive Bayes. Hasil yang diperoleh adalah 77.65% akurasi, 75% *precision*, 72% *recall*, dan 73% *f1-score*. Hal ini juga menyimpulkan bahwa penggunaan Complement Naive Bayes menghasilkan hasil yang lebih akurat dan tinggi untuk data yang tidak seimbang jika dibandingkan dengan penggunaan Multinomial Naive Bayes.

### 5.2 Saran

Berdasarkan penelitian yang telah dilakukan, berikut adalah beberapa saran yang dapat diambil untuk penelitian atau pengembangan lanjutan.

1. Membuat kamus yang mengandung banyak *slang words* dalam bahasa Indonesia dan umum digunakan oleh masyarakat pada media sosial. Dengan adanya kamus ini, maka proses *preprocessing* khususnya tahap normalisasi, *stemming*, dan *labeling* dapat menghasilkan data yang lebih baik dan menghasilkan model yang lebih akurat.

2. Mencoba menambahkan proses seperti validasi dengan menggunakan *cross validation* dan juga proses *translate* pada tahap *preprocessing* dengan menggunakan *library* Googletrans untuk kalimat yang memiliki lebih dari satu bahasa.
3. Mencoba menggunakan kamus atau *library* lain yang menghasilkan akurasi lebih baik dari *lexicon inset* untuk melakukan *labeling* bahasa Indonesia, misalkan dengan menggunakan *sentiword.id*.

