

## BAB V PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil dari analisis sentimen yang dilakukan pada *tweet* mengenai kebocoran data sim menggunakan algoritma *Naïve Bayes* dapat disimpulkan sebagai berikut:

1. Berdasarkan hasil analisis sentimen yang dilakukan pada 901 data *tweet* yang sudah diberi label, dapat diketahui bahwa klasifikasi dengan menggunakan algoritma *multinomial naïve bayes* dengan rasio pembagian data 80:20 dan dilakukan *oversampling* SMOTE pada data *train* mendapatkan hasil klasifikasi dengan nilai *accuracy* sebesar 71%, *precision* sebesar 62%, *recall* sebesar 74%, dan *f1-score* sebesar 62%. Kemudian hasil dari validasi menggunakan *k-fold cross validation* dengan *5-fold* didapatkan nilai rata-rata akurasi sebesar 72.13%.
2. Dari 901 data *tweet* yang digunakan dalam penelitian mengenai kebocoran data registrasi sim selama bulan September 2022. Frekuensi penggunaan kata yang paling sering muncul adalah “duga”, “bocor”, “data”, “kartu”, “sim”, “card”, “kominfo” dan diikuti oleh kata-kata lainnya. Dapat disimpulkan bahwa kata-kata tersebut berpengaruh bagi ketiga kelas sentimen ditandai dengan munculnya kata pada ketiga kelas sentimen. Sentimen negatif menjadi kelas sentimen terbanyak pada penelitian ini dengan jumlah sentimen yaitu 524 *tweet*.

### 5.2 Saran

Setelah didapatkan kesimpulan, masih terdapat beberapa kekurangan sehingga penulis berharap penelitian ini dapat dikembangkan dan mendapatkan hasil lebih baik. Adapun beberapa saran adalah sebagai berikut:

1. Melakukan *preprocessing* normalisasi dan *stopword* kata pada dokumen *tweet* dengan algoritma yang lebih baik dan *list* kata yang lebih tepat agar

tidak ada kata yang memiliki pengaruh terhadap dokumen terbuang.

2. Melakukan validasi terhadap dokumen yang telah diberi label dengan *InSet Lexicon* agar mendapatkan nilai sentimen yang akurat dan sesuai dengan bahasa yang digunakan. Ataupun bisa melakukan pelabelan dokumen secara manual dengan bantuan ahli bahasa.
3. Menambahkan algoritma klasifikasi teks lain untuk membandingkan tingkat akurasi hasil klasifikasi dari tiap algoritma.

