

BAB V PENUTUP

5.1 Kesimpulan

1. Data yang digunakan pada penelitian ini berupa teks *tweet* dari *twitter* dan didapatkan dengan *crawling* menggunakan pemrograman R. *Crawling* dilakukan sebanyak 5 kali pada tanggal 3, 10, 17, 24, 31 Juli 2021 menggunakan kata kunci #ppkmdarurat dan didapatkan sebanyak 5.000 *tweet*. Setelah data melalui tahapan *filtering* dan pelabelan, data menjadi 770 dengan 50% *tweet* berlabel ‘positif’ dan 50% *tweet* berlabel ‘negatif’. Data akan melalui tahapan *preprocessing* yang terdiri dari 6 tahapan yaitu *cleaning*, *case folding*, normalisasi bahasa, *stopword removal*, *stemming* dan tokenisasi. Data akan melalui tahapan pembobotan *term* (TF-IDF) bertujuan untuk memberikan bobot pada setiap kata. Selanjutnya data akan melalui tahapan seleksi fitur *Information Gain* dengan tujuan untuk meningkatkan performa metode *naive bayes*. Setelah melalui klasifikasi *Naïve Bayes* dan dievaluasi menggunakan metode *confusion matrix*.
2. Dari hasil penelitian menunjukkan adanya peningkatan model *Naïve Bayes* apabila menggunakan seleksi fitur *Information Gain* dengan nilai pengambilan *top ranking* ‘>0.0001’ yaitu akurasi 0.81, *recall* 0.82, *precision* 0.84, *f1 score* 0.83 dan *specificity* 0.79 dibandingkan sebelumnya yaitu akurasi 0.79, *recall* 0.81, *precision* 0.81, *f1 score* 0.81 dan *specificity* 0.76.

5.2 Saran

1. Diharapkan untuk penelitian selanjutnya dapat lebih mengoptimalkan proses *preprocessing*, seperti menambah *corpus* normalisasi bahasa.
2. Diharapkan untuk penelitian selanjutnya dapat menggunakan metode seleksi fitur lainnya untuk meningkatkan performa algoritma *naïve bayes*.
3. Diharapkan untuk penelitian selanjutnya dapat mencoba algoritma analisis sentimen yang lainnya seperti *Support Vector Machine* (SVM) atau *K-Nearest Neighbor* (K-NN).