

BAB V

KESIMPULAN

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan terhadap analisis sentimen kebijakan kenaikan harga BBM di Indonesia, dapat disimpulkan sebagai berikut ini:

1. Algoritma *Multinomial Naïve Bayes* dan *Random Forest* berhasil diimplementasikan untuk klasifikasi sentimen Twitter terkait kebijakan kenaikan harga BBM di Indonesia dengan menggunakan ratusan ribu data.
2. Pada tahap *preprocessing* data yaitu pelabelan yang merupakan implementasi metode *lexicon-based* dalam sentimen analisis. Diketahui bahwa secara keseluruhan masyarakat Indonesia cenderung memiliki sentimen netral terhadap permasalahan kebijakan kenaikan harga BBM dengan akumulasi jumlah data sebesar 136.577 *tweet*, sedangkan sentimen negatif dan positif masing-masing sebesar 76.386 *tweet* dan 34.495 *tweet*.
3. Hasil pengujian *tuning hyperparameter* pada *Naïve Bayes* didapatkan bahwa nilai terbaik parameter *alpha* atau nilai *laplace smoothing* adalah 5, agar mendapatkan nilai akurasi yang tinggi untuk model NBC.
4. Hasil pengujian *tuning hyperparameter* pada *Random Forest* didapatkan bahwa nilai *minimum samples split* sebesar 1 dan *minimum samples leaf* sebesar 1 dapat memaksimalkan hasil ketepatan klasifikasi. Berdasarkan skenario pengujian, diketahui juga bahwa parameter *min samples leaf* memberikan kenaikan akurasi paling signifikan hingga mencapai 5%.
5. Implementasi *5-fold cross validation* berhasil diimplementasi pada masing-masing model NBC dan RF. Dapat disimpulkan bahwa kedua model memiliki konsistensi nilai ketepatan klasifikasi yang baik, karena selisih nilai dari hasil *cross validation* dan tanpa *cross validation* tidak jauh berbeda. Dapat diartikan juga bahwa kedua model bebas dari kemungkinan *overfitting*.
6. Besar akurasi model terbaik *Naïve Bayes* dari hasil *tuning hyperparameter* yaitu 79.74%, lalu untuk model terbaik *Random Forest* sebesar 85.15%. Didapatkan juga nilai rata-rata AUC-ROC pada NBC sebesar 89.83% dan RF sebesar 94.62%. Jadi dari nilai keseluruhan, algoritma *Random Forest*

memiliki performa ketepatan klasifikasi yang lebih baik untuk mengklasifikasi teks dalam studi kasus analisis sentimen masyarakat Indonesia terhadap kebijakan kenaikan harga BBM.

5.2. Saran

Adapun beberapa saran yang dapat diberikan pada penelitian ini agar menjadi acuan dalam pengembang lebih lanjut adalah sebagai berikut:

1. Pada dataset teks bahasa Indonesia memiliki variasi kata yang terlalu besar, sehingga *dataset spelling normalization* pada tahap *preprocessing* masih perlu tambahkan kamus kata.
2. Dapat menambahkan metode pengolahan fitur tambahan seperti seleksi fitur menggunakan *Chi-square*.
3. Dapat menerapkan metode SMOTE sebagai cara mengatasi *dataset* yang *imbalance*.
4. Dapat menggunakan algoritma lain yang relevan untuk analisis sentimen seperti *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)* dan *Logistic Regression* untuk dilakukan analisis performansi.