

BAB 5

SIMPULAN DAN SARAN

5.1. Simpulan

Berdasarkan uji coba dan analisis yang telah dilakukan pada penelitian ini, dapat disimpulkan bahwa:

1. Implementasi algoritma *Multinomial Naïve Bayes* untuk menganalisa sentimen mengenai vaksinasi COVID-19 telah berhasil dibuat dengan membandingkan model pembobotan TF-IDF *Vectorizer* dan *Count Vectorizer*. Data yang digunakan untuk penelitian ini adalah data hasil *crawling* dari *Twitter* yang kemudian pemberian labelnya dilakukan secara manual oleh tiga orang dan diambil kesimpulan labelnya berdasarkan suara label terbanyak. Jumlah *tweet* yang didapatkan setelah melalui *text pre-processing* adalah 4141 *tweet* dengan rincian 2255 *tweet* berlabel netral, 1799 *tweet* berlabel positif, dan 87 *tweet* yang berlabel negatif.
2. Berdasarkan hasil uji coba pada data yang tidak seimbang, implementasi algoritma *Multinomial Naïve Bayes* dengan menggunakan model pembobotan *Count Vectorizer* dapat mengklasifikasi lebih baik dibandingkan TF-IDF *Vectorizer* jika pada penelitian memperhatikan hasil prediksi yang diklasifikasi benar dari masing-masing label. Dari hasil validasi menggunakan *Stratified 10-Fold Cross Validation*, *F1-Score* yang didapatkan tertinggi adalah 55,01% dengan $min_df = 5$ pada *Count Vectorizer*. Sedangkan, jika memperhatikan akurasi dari jumlah data yang diklasifikasi benar, maka *Multinomial Naïve Bayes* dengan model pembobotan TF-IDF *Vectorizer* yang lebih baik, dengan nilai *accuracy* yang mencapai 71,19% pada saat $min_df = 1$. Lalu, untuk pengujian implementasi algoritma *Multinomial Naïve Bayes* pada data yang seimbang (5397 *tweet*) yang di-*resample* sebelum *data splitting* memberikan hasil lebih baik dibandingkan data yang tidak seimbang. Klasifikasi terbaik pada saat menggunakan data seimbang diperoleh pada saat menggunakan model pembobotan TF-IDF *Vectorizer* dengan $min_df = 1$, baik dari metrik *accuracy* dan *f1-score* dengan rincian *accuracy* sebesar 83,18% dan *f1-score* sebesar 82,81%.

5.2. Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran yang dapat penulis berikan untuk pengembangan selanjutnya, antara lain:

1. Melihat dari penggunaan algoritma *Multinomial Naïve Bayes* dengan bantuan model pembobotan kata *TF-IDF Vectorizer*, akurasi *TF-IDF Vectorizer* dapat ditingkatkan melebihi *Count Vectorizer* jika menggunakan data yang seimbang dan mencoba memiliki data yang memiliki kata lebih bervariasi.
2. Mencoba menggunakan model pembobotan kata yang lainnya, seperti *Word2Vec* yang dapat mempelajari kata-kata yang memiliki sinonim atau menggunakan model *BERT (Bidirectional Encoder Representations from Transformers)* yang dapat membedakan arti dari kata yang ada sehingga dapat meningkatkan performa klasifikasi pada data yang seimbang maupun tidak seimbang.

