

BAB VI PENUTUP

Bab ini menjelaskan tentang kesimpulan dan saran dari penelitian yang telah dilakukan.

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dalam melakukan analisis sentimen terhadap pendapat masyarakat tentang program vaksinasi COVID-19 pada media sosial Twitter menggunakan Word2Vec dan *machine learning*, dapat diambil simpulan sebagai berikut :

1. Aplikasi untuk menganalisis sentimen data tweet mengenai pendapat masyarakat terhadap vaksinasi COVID-19 sudah dianalisis dan dibangun. Penelitian ini telah mengumpulkan data 5.355 data tweet yang diambil dari media sosial Twitter. Data tweet memiliki sentimen positif, dan negatif. Selain itu, aplikasi sudah dirancang desainnya dengan membuat UML. UML yang telah dirancang, yaitu *use case diagram*, *activity diagram*, dan *sequence diagram*.
2. Aplikasi dapat digunakan untuk memproses data tweet, menganalisis sentimen data tweet, dan mengklasifikasi sentimen data tweet. Dilakukan *preprocessing* terhadap data tweet untuk membersihkan data yang digunakan. Tahap *preprocessing* data yang terdiri dari *case folding*, *noise removal*, normalisasi, *stopword removal*, dan *stemming*.
3. Model Word2Vec menggunakan data 2 label telah dibangun untuk mempersiapkan data tweet berisi sentimen masyarakat terhadap vaksinasi COVID-19 untuk digunakan saat pemodelan. Model ini dapat digunakan untuk menghasilkan representasi vektor dari kata-kata dalam tweet. Word2Vec menggunakan model dengan parameter yang terbaik dari hasil analisis pengujian *hyperparameter tuning* yang telah dilakukan. Parameter model Word2Vec dibangun dengan ukuran dimensi 100 dan *window* 10 untuk *Support Vector Machine* (SVM), ukuran dimensi 200 dan *window* 7 untuk *Extreme Gradient Boosting* (XGBoost), dan ukuran dimensi 50 dan *window* 10 untuk Naive Bayes.

4. Algoritma *machine learning* *Support Vector Machine* (SVM), *Extreme Gradient Boosting* (XGBoost), dan Naive Bayes telah berhasil diuji menggunakan model klasifikasi untuk mengklasifikasi sentimen data tweet masyarakat terhadap vaksinasi COVID-19. Model klasifikasi yang memiliki kinerja yang baik memiliki *f1-score* dan akurasi yang tinggi dan kelas data yang seimbang. Berdasarkan hasil percobaan, SVM dengan parameter dimensi 50 dan *window* 10 menghasilkan *f1-score* sebesar 74.9% dan akurasi sebesar 75.1%, XGBoost dengan parameter dimensi 200 dan *window* 7 menghasilkan *f1-score* sebesar 90.3% dan akurasi sebesar 90.3%, dan Naive Bayes dengan parameter dimensi 50 dan *window* 10 menghasilkan *f1-score* sebesar 69.5% dan akurasi sebesar 69.7%. Hasil uji model menunjukkan bahwa model XGBoost memberikan hasil terbaik dengan *f1-score* dan nilai akurasi tertinggi dan kelas data yang seimbang.
5. Aplikasi *web* telah dibangun menggunakan bahasa pemrograman Python dengan *framework* *Flask* dan diimplementasikan model XGBoost sebagai model klasifikasi, model *Word2Vec* dengan ukuran dimensi 200 dan *window* 7, dan *SMOTE* untuk mengatasi *imbalance* data. Aplikasi ini dapat digunakan untuk menganalisis dan mengklasifikasi sentimen data tweet terhadap pendapat masyarakat mengenai vaksinasi COVID-19 dengan kategori sentimen positif maupun negatif.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat saran yang dapat diimplementasikan untuk pengembangan selanjutnya, diantaranya sebagai berikut:

1. Pertimbangan untuk menambahkan jumlah data yang digunakan untuk pelatihan data agar model klasifikasi yang dihasilkan lebih baik. Dengan *dataset* yang lebih besar, model klasifikasi dapat belajar pola-pola yang lebih kompleks dan mungkin menghasilkan performa yang lebih baik dalam mengklasifikasikan sentimen.
2. Meskipun XGBoost telah terbukti sebagai model klasifikasi terbaik dalam penelitian ini, dapat dilakukan penggunaan model lainnya untuk meningkatkan akurasi dan *f1-score* lebih lanjut. Contohnya teknik *deep*

learning, seperti model Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), dan transfer learning seperti BERT, GPT-3.

