BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut. Pertama, analisis distribusi sentimen terhadap 9.589 komentar YouTube yang diperoleh melalui YouTube DATA API v3 dan telah melalui tahapan *preprocessing* menunjukkan bahwa komentar masyarakat terkait Program Makan Bergizi Gratis didominasi oleh sentimen negatif sebanyak 5.039 komentar (52,5%), diikuti sentimen netral sebanyak 2.517 komentar (26,3%), dan sentimen positif sebanyak 2.032 komentar (21,2%). Pelabelan sentimen dilakukan secara otomatis menggunakan model IndoBERT (w11wo/indonesian-roberta-base-sentiment-classifier) sebagai pendekatan *pseudo-labelling*.

Kedua, algoritma *Naive Bayes* dengan teknik *Easy Data Augmentation* (EDA) menggunakan strategi *Random Word Swap* dan pelabelan IndoBERT berhasil diterapkan dalam proses klasifikasi sentimen. Teknik EDA efektif mengatasi ketidakseimbangan distribusi kelas dengan menyeimbangkan data menjadi 5.039 per kelas. Ekstraksi fitur menggunakan TF-IDF dengan konfigurasi N-Gram (1,2) dan optimasi hyperparameter melalui GridSearchCV menghasilkan model dengan alpha=0,1 sebagai konfigurasi optimal.

Ketiga, model *Naive Bayes* mencapai *accuracy* keseluruhan sebesar 84,06% dengan *F1-Score* per kelas: negatif (0,81), netral (0,83), dan positif (0,88). Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang cukup baik dan konsisten dalam mengklasifikasikan sentimen komentar pengguna, dengan

performa terbaik pada kelas positif (recall 0,92) dan keseimbangan yang baik antar kelas setelah penerapan teknik augmentasi data.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang telah diidentifikasi, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya :

Pertama, penelitian selanjutnya dapat mencoba menggunakan algoritma klasifikasi lain sebagai pembanding, seperti *Support Vector Machine (SVM), Random Forest, Logistic Regression* atau metode berbasis *deep learning* seperti LSTM dan BERT fine-tuning, untuk menentukan metode yang paling optimal dalam analisis sentimen pada topik yang sama. Perbandingan performa antar algoritma dapat memberikan *insight* yang lebih komprehensif mengenai kecocokan masing-masing metode dengan karakteristik data sentimen berbahasa Indonesia.

Kedua, dapat dilakukan eksplorasi lebih lanjut mengenai berbagai strategi augmentasi teks lainnya dari teknik EDA, seperti *Synonym Replacement* atau *Random Insertion*, serta membandingkan efektivitasnya dengan teknik *oversampling* berbasis matriks seperti SMOTE atau metode penyeimbangan data lainnya. Penentuan parameter augmentasi (aug_p) yang optimal melalui eksperimen sistematis juga dapat meningkatkan kualitas data sintetis yang dihasilkan.

Ketiga, penelitian selanjutnya dapat memperluas sumber data dengan mengambil data dari periode waktu yang lebih panjang, menambahkan variasi kata kunci atau menggabungkan data dari platform media sosial lainnya seperti X

(Twitter), Instagram atau TikTok, sehingga kajian yang dilakukan lebih menyeluruh dan representatif terhadap opini publik secara keseluruhan.

Keempat, proses pelabelan sentimen dapat ditingkatkan dengan menggunakan lebih dari satu model untuk pelabelan ensemble, atau melakukan validasi pada sampel data yang dilabeli secara manual oleh ahli bahasa untuk mengukur tingkat kesepakatan *(inter-annotator agreement)* antara label model dengan label manusia, sehingga reliabilitas hasil pelabelan *pseudo-labelling* dapat terukur secara lebih akurat.

Kelima, analisis lebih mendalam terhadap data yang salah diklasifikasikan *(misclassification analysis)* dapat dilakukan untuk mengidentifikasi pola kesalahan model, sehingga dapat dikembangkan strategi perbaikan yang lebih tepat sasaran, baik dari sisi *preprocessing*, ekstraksi fitur maupun pemilihan algoritma klasifikasi.