

## BAB 5

### SIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, berikut adalah beberapa hal yang dapat disimpulkan:

1. Implementasi analisis sentimen menggunakan algoritma IndoBERTweet untuk mengklasifikasikan *tweet* berbahasa Indonesia terkait Program Makan Bergizi Gratis berhasil dilakukan. Data dikumpulkan melalui *crawling* Platform X menghasilkan 17.480 *tweet*, melalui prapemrosesan dan pelabelan dengan dua pendekatan: *heuristic labeling* (17.480 *tweet*) dan *pseudo-labeling* (10.498 *tweet*) dengan *threshold confidence*  $\geq 70\%$ .
2. Nilai akurasi terbaik sebesar 97,67% diperoleh dari konfigurasi *Pseudo-labeling* dengan *preprocessing Default* pada data *testing* (1.575 sampel). Model menunjukkan performa seimbang dengan *Precision* 97,76%, *Recall* 97,67%, dan *F1-Score* 97,70%. Analisis *Confusion Matrix* menunjukkan akurasi per kelas: Negatif 98,27%, Netral 93,18%, dan Positif 98,03%, dengan total prediksi benar 1.539 dari 1.575 sampel. Kelas Netral menunjukkan tingkat kesalahan lebih tinggi (6,82%) dibandingkan kelas ekstrem, mengindikasikan kesulitan pembedaan dengan kelas lain.
3. Perbandingan kedua strategi pelabelan dengan konfigurasi hyperparameter sama menunjukkan hasil berbeda menurut jenis *preprocessing*. Pada *preprocessing Paper*, *Heuristic* menghasilkan *F1-Score* 96,10% lebih baik daripada *Pseudo-labeling* 94,92% (+1,17%). Sebaliknya, pada *preprocessing Default*, *Pseudo-labeling* mencapai 97,70% unggul dari *Heuristic* 92,00% (+5,70%). Keunggulan *Pseudo-labeling* pada *preprocessing Default* disebabkan kualitas label lebih tinggi melalui *seed* manual dan *confidence threshold*, distribusi data lebih seimbang (Negatif 19,12%, Netral 32,74%, Positif 48,14%), dan stabilitas lebih baik dengan *gap* validasi *testing* lebih kecil. Hasil ini menunjukkan kombinasi *pseudo-labeling* dan *preprocessing* komprehensif meningkatkan signifikan performa IndoBERTweet dalam klasifikasi sentimen *tweet* berbahasa Indonesia.

## 5.2 Saran

Berdasarkan temuan penelitian, disarankan agar penelitian lanjutan:

1. Mengeksplorasi konfigurasi pelatihan lebih luas (mis. jumlah *epoch* lebih dari 3, *learning rate*, dan *early stopping*) untuk menemukan durasi pelatihan yang optimal.
2. Melakukan beberapa run dengan *seed* berbeda dan melaporkan mean serta standar deviasi metrik (mis. *F1-score*) untuk mengukur stabilitas hasil.
3. Uji model transformer lain misalnya (IndoBERT, mBERT, XLM-R) untuk melihat kombinasi model apakah meningkat hasilnya.

