

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada Bab IV, penelitian mengenai analisis sentimen ulasan aplikasi TikTok di Google Play Store menggunakan arsitektur hibrid IndoBERT dan Convolutional Neural Network (CNN) 1-Dimensi ini dapat disimpulkan sebagai berikut:

1. Arsitektur hibrid IndoBERT dan 1D-CNN berhasil diimplementasikan secara utuh untuk melakukan analisis sentimen pada ulasan aplikasi TikTok berbahasa Indonesia, menjawab rumusan masalah pertama penelitian ini. IndoBERT berperan sebagai lapisan ekstraksi representasi semantik kontekstual berdimensi 768, sedangkan lapisan 1D-CNN (Conv1D, ReLU, Global Max Pooling, dan Dropout) berperan mengekstraksi fitur lokal n-gram sebelum diklasifikasikan oleh Fully Connected Layer ke dalam tiga kelas sentimen. Model yang dilatih selama 5 epoch menunjukkan proses pembelajaran yang stabil, ditandai dengan penurunan training loss dari 0,6992 menjadi 0,2563 dan peningkatan training accuracy dari 72,11% menjadi 92,96%. Namun demikian, ketika diuji pada 89 data uji yang belum pernah dilihat model, performa generalisasi model tercatat pada tingkat akurasi keseluruhan sebesar **85,39%**, dengan *weighted average* F1-score 0,83. Model menunjukkan performa yang baik dalam mengklasifikasikan kelas mayoritas, yaitu Positif (F1-score 0,92) dan Negatif (F1-score 0,64), tetapi mengalami kegagalan total dalam mengenali kelas Netral (F1-score 0,00, tidak ada satu pun data Netral yang berhasil diklasifikasikan dengan benar). Kesenjangan antara *macro average* F1-score (0,52) dan *weighted average* F1-score (0,83) mengonfirmasi bahwa performa model secara keseluruhan sangat dipengaruhi oleh dominasi kelas mayoritas, sementara kemampuan model dalam menangani kelas minoritas masih jauh dari memadai. Tahapan prapemrosesan teks (*case folding*, *cleansing*, normalisasi, dan *stopword removal*) terbukti efektif dalam menangani sebagian karakteristik *noise* pada data ulasan, menjawab rumusan masalah kedua

penelitian ini. Pipa prapemrosesan berhasil menyeragamkan format huruf, membersihkan simbol dan emoji, mengubah sebagian kata tidak baku/singkatan menjadi bentuk baku, serta membuang kata-kata fungsional yang tidak bermuatan sentimen, sehingga menyisakan 444 data ulasan bersih dari 482 data mentah. Meskipun demikian, hasil pengamatan pada Tabel 4.2.3 menunjukkan bahwa kamus normalisasi yang digunakan belum mencakup seluruh variasi kata tidak baku yang muncul pada dataset (seperti "tantal", "tt", "dri", "jdi", dan "pdahal"), sehingga efektivitas tahap normalisasi masih dapat ditingkatkan lebih lanjut.

2. Ketidakseimbangan kelas (*class imbalance*) pada dataset di mana kelas Netral hanya menyumbang sekitar 5,3% dari keseluruhan data teridentifikasi sebagai faktor utama yang membatasi performa model secara keseluruhan, khususnya dalam mengenali kelas Netral. Hal ini menunjukkan bahwa akurasi tinggi pada tingkat agregat tidak serta-merta mencerminkan kemampuan model yang seimbang dalam mengklasifikasikan seluruh kelas sentimen secara adil.

5.2 Saran

Berdasarkan kesimpulan di atas, beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya maupun pengembangan lebih lanjut terhadap penelitian ini antara lain:

1. Penanganan Ketidakseimbangan Data (*Class Imbalance*): Penelitian selanjutnya disarankan menerapkan teknik penanganan ketidakseimbangan kelas secara eksplisit, seperti pembobotan kelas pada fungsi kerugian (*class-weighted loss*), teknik *oversampling* pada data teks (misalnya melalui augmentasi data seperti *back-translation* atau *synonym replacement*), maupun pengumpulan data tambahan khusus kelas Netral, guna meningkatkan kemampuan model dalam mengenali kelas minoritas.
2. Perluasan Kamus Normalisasi: Cakupan kamus normalisasi kata tidak baku (*slang dictionary*) perlu diperluas secara berkelanjutan berdasarkan hasil observasi pada dataset yang lebih besar, agar proses penyelarasan kata gaul menjadi bentuk baku dapat berjalan lebih optimal sebelum diproses oleh IndoBERT.

3. Validasi Label secara Manual: Mengingat proses pelabelan pada penelitian ini bersifat otomatis berdasarkan skor rating (*distant supervision*), penelitian selanjutnya disarankan melakukan validasi manual pada sebagian sampel data, khususnya pada ulasan berbintang 3, karena rating tersebut tidak selalu secara konsisten mencerminkan sentimen netral pada teksnya.
4. Studi Ablasi dan Perbandingan Arsitektur: Sebagaimana disebutkan pada sub-bab 3.1.1, perbandingan performa antara model hibrid IndoBERT + 1D-CNN dengan model *baseline* IndoBERT murni (tanpa lapisan CNN) perlu dilengkapi untuk membuktikan secara empiris kontribusi nyata lapisan 1D-CNN terhadap peningkatan performa klasifikasi.
5. Eksplorasi Hyperparameter Lanjutan: Penelitian selanjutnya dapat mengeksplorasi penambahan jumlah epoch disertai teknik *early stopping*, penjadwalan *learning rate* (*learning rate scheduler*), maupun pengaturan ulang nilai *dropout rate*, untuk melihat potensi peningkatan performa model tanpa menimbulkan *overfitting*.
6. Perluasan Sumber dan Volume Data: Penelitian mendatang dapat mempertimbangkan penambahan sumber data dari platform lain (seperti Apple App Store atau media sosial) serta volume data yang lebih besar, khususnya untuk memperbanyak jumlah sampel pada kelas Netral yang selama ini menjadi keterbatasan utama.
7. Reprodutifitas dan Snapshot Dataset: Mengingat proses pengambilan data dilakukan secara dinamis (real-time scraping) berdasarkan ulasan terbaru, hasil eksperimen yang diperoleh bersifat spesifik terhadap rentang waktu pengambilan data tersebut (Januari 2026 - Juni 2026) dan berpotensi menghasilkan angka yang sedikit berbeda apabila proses scraping diulang pada waktu yang berbeda. Penelitian selanjutnya disarankan untuk menyimpan dan mempublikasikan salinan tetap (snapshot) dataset yang digunakan, guna menjamin reprodutifitas dan validasi ulang hasil penelitian oleh pihak lain.