

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil eksperimen dan evaluasi yang telah dilakukan terhadap tiga model Transformer, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa pada tugas klasifikasi emosi teks Bahasa Indonesia, dapat ditarik kesimpulan sebagai berikut:

1. Ketiga model, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa menunjukkan performa yang baik dalam tugas klasifikasi emosi. Hal ini ditunjukkan dengan nilai akurasi, presisi, *recall*, dan *macro f1-Score* yang relatif tinggi pada seluruh skema pembagian data. Hal ini juga menjadi pembuktian bahwa pendekatan berbasis Transformer efektif untuk memahami konteks emosional dalam teks Bahasa Indonesia.
2. Terdapat perbedaan waktu komputasi antar model. Meskipun seluruh model memiliki arsitektur berbasis Transformer dengan jumlah parameter besar, terdapat variasi dalam waktu pelatihan. Waktu komputasi berada pada rentang 2 hingga 4 menit per skema, dengan IndoBERT sebagai model dengan waktu komputasi tercepat pada skema *holdout* dan IndoRoBERTa sebagai model dengan waktu pelatihan paling cepat pada skema *k-fold*.
3. Berdasarkan hasil evaluasi keseluruhan, IndoBERTweet menunjukkan performa terbaik dalam tugas klasifikasi emosi berbasis teks Bahasa Indonesia. Berdasarkan rata-rata hasil K-fold cross validation, IndoBERTweet memperoleh nilai macro F1-score tertinggi yaitu sebesar 0,7807 pada skema K=5 dan 0,7778 pada skema K=6, sehingga menunjukkan performa yang paling konsisten dibandingkan model lainnya. Selain itu, pada skema terbaik yaitu K-fold dengan nilai k = 5 pada fold ke-2, IndoBERTweet memperoleh nilai accuracy sebesar 80,3%, precision sebesar 80,47%, recall sebesar 81,89%, dan macro F1-score sebesar 80,99%, yang merupakan nilai tertinggi pada seluruh

eksperimen.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemui selama proses eksperimen, beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Eksplorasi *hyperparameter* secara lebih mendalam. Penelitian ini menggunakan konfigurasi pelatihan standar tanpa melakukan eksplorasi *hyperparameter*. Penelitian selanjutnya dapat menerapkan teknik optimasi seperti Grid Search, Random Search, atau Bayesian Optimization untuk memperoleh kombinasi parameter terbaik yang berpotensi meningkatkan performa model.
2. Penanganan ketidakseimbangan data dalam *dataset*. Distribusi data antar kelas dalam penelitian ini tidak sepenuhnya seimbang. Penelitian selanjutnya dapat menerapkan teknik seperti *class weighting*, *oversampling*, atau metode *data augmentation* untuk meningkatkan kemampuan model dalam mengenali kelas minoritas.
3. Analisis *attention visualization* atau teknik interpretasi lainnya untuk memahami bagaimana model mengambil keputusan dalam menentukan label emosi.
4. Pengujian pada *dataset* yang lebih beragam untuk menguji kemampuan generalisasi model. Penelitian selanjutnya dapat menggunakan *dataset* dengan domain berbeda seperti berita, forum diskusi, atau komentar produk, sehingga dapat diketahui ketahanan model pada variasi konteks bahasa yang lebih luas.
5. Pengembangan implementasi *real-time* karena implementasi pada penelitian ini masih bersifat sederhana dan berjalan secara lokal. Penelitian selanjutnya dapat mengembangkan sistem berbasis *web* atau API yang dapat diakses secara *online* untuk mendukung aplikasi analisis emosi secara *real-time*.