

BAB V

SIMPULAN SARAN

5.1 Simpulan

Penelitian ini berhasil melakukan komparasi tiga model *Transformer* BERT, RoBERTa, dan FNet untuk analisis sentimen berbasis teks ulasan produk *fashion*. Hasil pengujian menunjukkan bahwa RoBERTa memberikan kinerja terbaik dengan akurasi 89,67%, disusul oleh BERT dengan 89,10%, dan FNet dengan 88,15%. RoBERTa unggul dalam menangkap konteks dan makna semantik yang lebih mendalam berkat strategi pelatihan yang dioptimalkan, sementara FNet menawarkan efisiensi komputasi yang lebih baik dengan waktu pemrosesan tercepat. Temuan ini menegaskan bahwa model berbasis *Transformer* mampu memberikan hasil yang akurat dan stabil untuk tugas klasifikasi sentimen teks.

Pada analisis sentimen berbasis gambar (*unimodal image analysis*), model CNN ResNet-18 menunjukkan performa yang bervariasi tergantung pada kualitas dan relevansi gambar yang digunakan. Hasil terbaik dicapai pada gambar sintetis yang dihasilkan oleh *SDXL Turbo*, dengan akurasi 75,18%, melampaui *Stable Diffusion 1.5* (63,50%) dan *v2-Base* (68,13%). Hal ini menunjukkan bahwa semakin baik kualitas dan kesesuaian visual gambar terhadap deskripsi produk, semakin tinggi pula kemampuan model dalam mengenali sentimen secara visual. ResNet-18 dipilih karena mampu memberikan keseimbangan optimal antara akurasi, waktu komputasi, dan stabilitas selama pelatihan.

Hasil perbandingan tiga model *Generative AI Stable Diffusion 1.5*, *Stable Diffusion v2-Base*, dan *SDXL Turbo* menunjukkan bahwa *SDXL Turbo* menghasilkan gambar sintetis paling realistik, dengan pencahayaan, tekstur, dan warna yang menyerupai citra produk asli. Berdasarkan preferensi manusia, gambar dari *SDXL Turbo* paling disukai karena menampilkan detail yang lebih halus dan kesan visual yang alami, sementara versi 1.5 dan *v2-Base* cenderung tampak lebih artifisial. Hal ini membuktikan bahwa *generative* model dengan kemampuan difusi

cepat dan distilasi tinggi seperti *SDXL Turbo* lebih efektif dalam melengkapi ulasan yang tidak memiliki gambar untuk mendukung analisis *multimodal*.

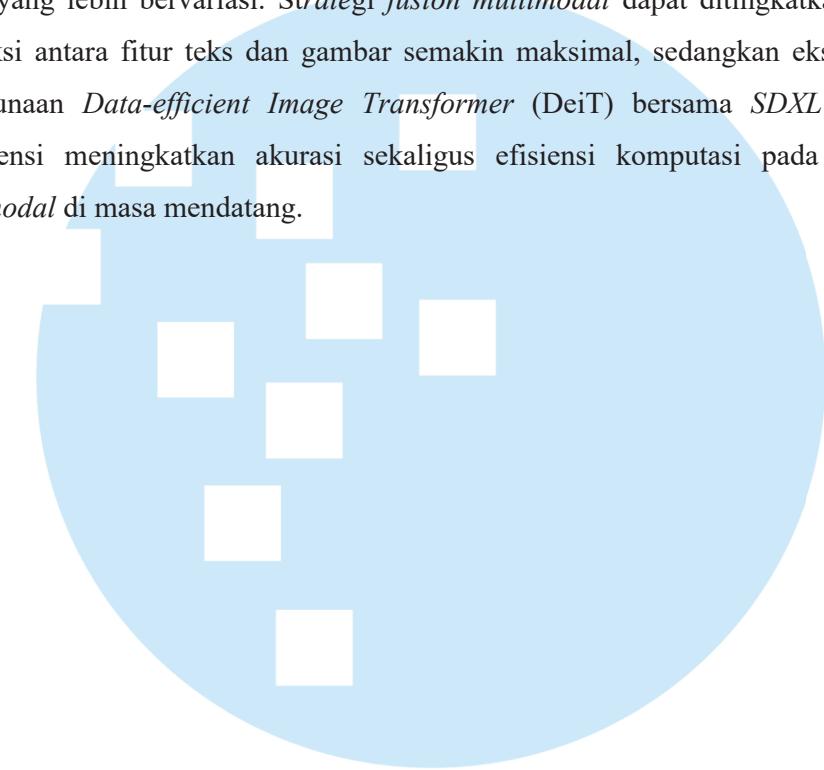
Dalam pengembangan dan pengujian *framework multimodal sentiment analysis* yang mengintegrasikan *Transformer*, CNN ResNet-18, dan *Generative AI*, diperoleh hasil bahwa model *SDXL Turbo*–ResNet-18–FNet memberikan performa terbaik dengan akurasi 91,20% dan waktu eksekusi 2,205 detik. Integrasi antara fitur teks dan gambar terbukti mampu menghasilkan representasi yang lebih kaya dibandingkan pendekatan *unimodal*, sekaligus meningkatkan nilai *precision*, *recall*, dan *F1-score* secara keseluruhan. Meskipun terdapat indikasi *overfitting* ringan akibat keterbatasan data dan repetisi citra sintetis, model *multimodal* ini menunjukkan potensi besar untuk diterapkan pada analisis sentimen produk *e-commerce*, terutama dalam kategori *fashion* yang mengandalkan konteks visual dan deskriptif secara bersamaan.

5.2 Saran

Penelitian ini masih memiliki beberapa keterbatasan yang perlu diperhatikan dalam pengembangan selanjutnya. Ukuran *dataset* yang relatif kecil serta dominasi data sintetis menyebabkan model berpotensi mengalami *overfitting* dan kurang mampu mengenali variasi data dunia nyata. Selain itu, sebagian foto ulasan pelanggan yang dikumpulkan dari *platform e-commerce* tidak selalu relevan dengan produk yang diulas. Beberapa gambar menampilkan objek acak, latar yang tidak sesuai, atau bahkan foto promosi toko alih-alih produk sebenarnya, sehingga menurunkan akurasi model dalam memahami keterkaitan antara teks dan gambar. Kualitas gambar hasil generatif juga belum sepenuhnya menyerupai kondisi produk nyata, sehingga masih terdapat kesenjangan antara representasi visual sintetis dan gambar dunia nyata.

Peningkatan kinerja model *multimodal* dapat dilakukan melalui penerapan teknik regularisasi seperti *dropout*, *weight decay*, atau *label smoothing* untuk mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model. Variasi data perlu diperluas dengan teknik data augmentation yang lebih beragam agar model mampu mengenali pola dari berbagai kondisi pencahayaan dan sudut pandang. Metode *object detection* dapat dimanfaatkan untuk memfilter gambar

yang relevan dari unggahan pelanggan sehingga hanya citra produk yang sesuai dan berkualitas baik yang digunakan dalam pelatihan. Pengembangan berikutnya juga dapat difokuskan pada perancangan *prompt* yang lebih optimal untuk menghasilkan gambar sintetis yang konsisten dan relevan, serta pada pemanfaatan *dataset* gambar nyata yang lebih bervariasi. Strategi *fusion multimodal* dapat ditingkatkan agar interaksi antara fitur teks dan gambar semakin maksimal, sedangkan eksplorasi penggunaan *Data-efficient Image Transformer* (DeiT) bersama *SDXL Turbo* berpotensi meningkatkan akurasi sekaligus efisiensi komputasi pada model *multimodal* di masa mendatang.



UMN