

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini berhasil meningkatkan *multimodal sentiment analysis* pada ulasan produk *e-commerce* melalui pemanfaatan *image filtering* dan *domain-adapted generative modeling* yang diintegrasikan dalam alur pemrosesan data dan pembentukan *dataset multimodal*, meskipun masih memiliki keterbatasan pada tahap *image filtering*. Salah satu komponen penting dalam alur tersebut adalah penyaringan awal gambar ulasan menggunakan YOLOv8, yang berperan menghapus gambar yang jelas tidak relevan, seperti foto kemasan, resi pengiriman, atau objek *non-fashion*, sehingga korpus visual menjadi lebih terfokus pada objek produk *fashion*. Namun, YOLOv8 tidak dirancang untuk membedakan kondisi pakaian yang rusak dan tidak rusak, sementara pada kelas sentimen negatif sebagian besar konsumen tidak selalu memotret bagian yang bermasalah dan sering kali hanya mengunggah foto produk secara umum atau paket. Akibatnya, hasil *filtering* otomatis belum mampu membentuk himpunan gambar negatif yang benar-benar mencerminkan keluhan pada teks ulasan, sehingga kurasi manual tetap diperlukan untuk memastikan representasi visual sentimen negatif sesuai dengan konteks keluhan konsumen. Temuan ini menegaskan bahwa YOLOv8 lebih tepat diposisikan sebagai alat bantu *pre-filtering* untuk relevansi objek, bukan sebagai satu-satunya mekanisme seleksi gambar berlabel sentimen.

Pendekatan *domain-adapted Generative AI* berbasis Stable Diffusion XL (SDXL) yang dioptimalkan dengan LoRA memungkinkan kekosongan modalitas visual khususnya pada ulasan negatif yang tidak menyertakan gambar dapat diatasi secara efektif. Melalui *fine-tuning* dengan pasangan gambar dan label kerusakan pakaian, model mampu menghasilkan 745 gambar sintetis negatif yang secara visual menonjolkan pola cacat seperti bahan tipis, jahitan lepas, atau bagian yang robek dan selaras dengan isi keluhan pada teks ulasan. Gambar sintetis ini melengkapi 255 gambar negatif nyata dan dipasangkan dengan 1.000 gambar

positif nyata, sehingga terbentuk *dataset multimodal* yang seimbang (1.000 positif dan 1.000 negatif) baik pada sisi teks maupun gambar. Keseluruhan hasil menunjukkan bahwa SDXL-LoRA berperan penting dalam memperkaya kanal visual ketika data dunia nyata terbatas, dengan catatan kualitas dan distribusi gambar generatif tetap perlu diawasi agar tidak menimbulkan bias baru pada proses pelatihan.

Analisis sentimen berbasis teks menunjukkan bahwa seluruh model *Transformer* yang diterapkan, yakni IndoBERT, BERT, dan RoBERTa, mampu mengklasifikasikan sentimen ulasan *fashion* dengan kinerja yang sangat tinggi, di mana IndoBERT secara konsisten menunjukkan keunggulan dibandingkan model lainnya. Capaian ini mengonfirmasi bahwa pemanfaatan model pralatih yang dibangun khusus untuk bahasa Indonesia memberikan keuntungan nyata dan lebih efektif dibandingkan pendekatan berbasis terjemahan ke bahasa Inggris. Meskipun demikian, model BERT dan RoBERTa tetap menunjukkan performa yang stabil serta waktu inferensi yang praktis, sehingga keluarga model *transformer* dalam penelitian ini layak dijadikan pondasi analisis sentimen, dengan IndoBERT sebagai opsi paling superior untuk konteks *e-commerce* lokal.

Analisis sentimen berbasis gambar memperlihatkan bahwa arsitektur visi modern mampu memberikan lompatan kinerja yang signifikan ketika dikombinasikan dengan pembersihan data yang ketat dan kurasi visual yang selaras dengan label sentimen. Model seperti DeiT Tiny dan ResNet-18 terbukti mampu mengungguli model *baseline* CNN konvensional dengan margin yang cukup jauh. Kinerja superior ini melampaui berbagai penelitian terdahulu yang kurang spesifik dalam memfiltrasi gambar, di mana himpunan gambar negatif dan positif seringkali tampak serupa secara visual. Hasil penelitian ini menegaskan vitalnya peran pembersihan data menyeluruh, baik dari sisi teks maupun gambar, termasuk pemilahan manual dan penambahan gambar negatif generatif, agar sinyal visual yang dipelajari oleh model benar-benar merefleksikan polaritas sentimen yang diwakili secara akurat.

Pendekatan *multimodal sentiment analysis* yang mengintegrasikan representasi teks dan gambar menghasilkan peningkatan kinerja yang nyata

dibandingkan pendekatan *unimodal*, sekaligus menyediakan konfigurasi model terbaik yang dapat dijadikan rujukan praktis. Kombinasi antara IndoBERT dan ResNet-18 berhasil memberikan performa tertinggi dengan kemampuan klasifikasi yang sangat seimbang antara ulasan positif dan negatif. Temuan ini diperkuat oleh hasil pemodelan topik yang mampu memetakan keluhan dominan konsumen terkait kualitas bahan dan kesesuaian produk, serta apresiasi terhadap harga dan pelayanan. Secara keseluruhan, hasil ini membuktikan bahwa fusi teks dan gambar pada *dataset* yang telah difilter menggunakan YOLOv8, dikurasi secara manual, dan diperkaya dengan gambar generatif SDXL-LoRA, berhasil menghasilkan model *multimodal* yang jauh lebih akurat, stabil, dan representatif, sehingga sangat siap untuk diadopsi dalam pemantauan sentimen konsumen pada ekosistem *e-commerce fashion* di Indonesia.

5.2 Saran

Berdasarkan hasil dan keterbatasan yang telah diuraikan, beberapa saran dapat diajukan untuk pengembangan penelitian selanjutnya. Pada sisi *image filtering*, disarankan untuk tidak hanya mengandalkan YOLOv8 sebagai penyaring relevansi objek, tetapi menambah model deteksi kerusakan pakaian yang lebih spesifik, misalnya melalui *fine-grained defect detection* atau klasifikasi atribut visual seperti “robek”, “bahan tipis”, dan “jahitan lepas” yang dilatih dari himpunan gambar beranotasi manual. Integrasi sinyal teks-gambar sejak tahap awal, misalnya dengan memanfaatkan *cross-modal attention* untuk menyesuaikan seleksi gambar negatif berdasarkan keluhan pada teks, juga berpotensi mengurangi kebutuhan kurasi manual yang intensif sambil menjaga keselarasan antara modalitas visual dan tekstual.

Pada pembentukan *dataset* dan rancangan model, penelitian selanjutnya disarankan untuk memperluas cakupan di luar kategori *fashion* dan di luar dua *platform e-commerce* yang digunakan, menambahkan kelas sentimen netral atau multi-aspek (misalnya aspek bahan, ukuran, pengiriman), serta meningkatkan variasi kualitas dan kondisi gambar agar model lebih robust terhadap konteks dunia nyata. Evaluasi terhadap gambar sintesis yang dihasilkan SDXL-LoRA sebaiknya dilengkapi dengan penilaian pakar maupun metrik kuantitatif untuk mengukur

realisme dan konsistensi dengan teks, serta dilakukan studi ablation untuk membandingkan kinerja model dengan dan tanpa data generatif guna memastikan bahwa penambahan gambar sintetis benar-benar meningkatkan performa tanpa menimbulkan *bias* baru.

Pada rancangan arsitektur dan penerapan sistem, penelitian berikutnya dapat mengeksplorasi model teks dan visi yang lebih mutakhir (misalnya varian *Transformer* yang lebih besar atau *vision transformer* generasi terbaru) serta skema fusi *multimodal* yang lebih kaya, seperti *cross-modal transformer* atau mekanisme perhatian hierarkis, bukan sekadar penggabungan vektor secara sederhana. Di sisi implementasi praktis, sistem *multimodal sentiment analysis* yang dihasilkan dapat dikembangkan menuju prototipe *dashboard* pemantauan sentimen bagi penjual maupun pengelola *platform*, dengan penekanan pada kemampuan adaptasi terhadap *drift* data (perubahan tren bahasa dan gaya foto) dan fasilitas *explainability* melalui visualisasi perhatian teks-gambar agar keputusan model lebih mudah dipahami dan dipercaya oleh pemangku kepentingan.

UMN