

BAB 5. PENUTUP

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan model *Visual Question Answering* (VQA) dengan menggunakan *backbone* teks BERT dan dua *backbone* visual berbeda, yaitu *Vision Transformer* (ViT) dan *EfficientNet V2*. Model diuji dengan dua strategi pelatihan, yaitu *transfer learning* dan *fine tuning* serta dua skenario *threshold* (jumlah minimal kemunculan kelas), yaitu 75 dan 50.

Hasil menunjukkan bahwa model dengan *backbone EfficientNet V2* dan BERT dengan metode *transfer learning* pada *threshold* 75 menghasilkan performa tertinggi, dengan *test accuracy* sebesar 56.30% dan *accuracyANLS* sebesar 62.90%. Sementara itu, model dengan *backbone* ViT dan BERT dengan *fine tuning* menunjukkan performa kompetitif, terutama pada *threshold* 75 (*accuracy* sebesar 53.19% dan *accuracyANLS* sebesar 61.95%), serta menunjukkan sensitivitas yang lebih tinggi terhadap variasi data pada *threshold* 50. Temuan ini menegaskan bahwa pemilihan *backbone* visual dan strategi pelatihan sangat memengaruhi performa model, serta pentingnya penyesuaian pendekatan berdasarkan karakteristik distribusi label dalam *dataset* VQA.

Sebagai bentuk penerapan model dalam konteks dunia nyata, penelitian ini juga mengembangkan sebuah aplikasi *Graphical User Interface* (GUI) berbasis desktop dengan menggunakan pustaka *TKInter*. Antarmuka ini memungkinkan pengguna untuk mengunggah gambar, mengajukan pertanyaan, dan menerima jawaban dari model secara langsung tanpa memerlukan akses internet atau keterampilan teknis khusus. Dengan demikian, GUI ini diharapkan dapat meningkatkan aksesibilitas dan kemudahan penggunaan model VQA.

5.2 Saran

Untuk pengembangan lebih lanjut, disarankan agar penelitian selanjutnya menggunakan *dataset* VQA dengan jumlah data yang lebih besar serta distribusi label yang lebih seimbang, guna menunjang pelatihan model berukuran besar secara optimal. Selain itu, eksplorasi terhadap arsitektur visual lainnya, seperti *SwIn Transformer*, versi *backbone* yang lebih kompleks, maupun penerapan teknik

pelatihan lanjutan seperti *self-supervised learning*, *reinforcement learning*, dan integrasi mekanisme OCR, dapat menjadi langkah strategis untuk meningkatkan akurasi model.

Dari sisi aplikasi, *Graphical User Interface* (GUI) yang telah dikembangkan pada platform desktop dapat diperluas ke platform lain, seperti aplikasi berbasis web atau *mobile*, guna menjangkau pengguna yang lebih luas. Penambahan fitur seperti pemrosesan *batch*, integrasi kamera secara langsung, serta penggunaan *input* suara (*speech-to-text*) juga dapat dipertimbangkan untuk meningkatkan fungsionalitas sistem. Selain itu, optimalisasi model agar dapat dijalankan pada perangkat dengan sumber daya terbatas, seperti *edge device* atau *embedded system*, akan sangat bermanfaat untuk mendukung penerapan VQA dalam berbagai skenario praktis.