

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil perancangan, implementasi, eksperimen, evaluasi, serta uji coba fungsional yang telah dilakukan, penelitian ini berhasil membangun sistem *image captioning* bilingual yang mengintegrasikan BLIP-2 sebagai penghasil *caption* Bahasa Inggris dan MarianMT sebagai penerjemah ke Bahasa Indonesia dalam satu *pipeline* terpadu. Sistem mampu menghasilkan deskripsi gambar secara otomatis dalam dua bahasa serta menyajikannya dalam bentuk teks dan audio melalui integrasi *Google Text-to-Speech* (gTTS) pada aplikasi web berbasis FastAPI. Integrasi visual, teks, dan suara tersebut membentuk sistem multimodal yang interaktif dan responsif, sehingga berpotensi mendukung peningkatan literasi visual siswa sekolah dasar.

Hasil evaluasi menggunakan metrik METEOR menunjukkan bahwa model BLIP-2 *pretrained* merupakan konfigurasi yang paling stabil untuk implementasi akhir, dengan skor 0,3765 pada Bahasa Inggris dan 0,3295 pada Bahasa Indonesia. Model QLoRA skema 80:10:10 memperoleh skor 0,3818 (Inggris) dan 0,3065 (Indonesia), sedangkan QLoRA 80:20 memperoleh skor 0,3696 (Inggris) dan 0,2946 (Indonesia). Meskipun *fine-tuning* QLoRA 80:10:10 menunjukkan peningkatan kecil pada Bahasa Inggris, BLIP-2 *pretrained* dipilih karena menghasilkan *caption* yang lebih stabil, konsisten, ringkas, serta minim halusinasi, sehingga lebih sesuai untuk kebutuhan deskripsi yang jelas dan terarah bagi siswa sekolah dasar.

Eksperimen *fine-tuning* juga menunjukkan bahwa penurunan *train loss* tidak selalu mencerminkan peningkatan kualitas semantik. Skema 80:20 memperlihatkan indikasi *overfitting*, di mana model menghasilkan *caption* yang lebih panjang dan naratif namun kurang selaras dengan visual, sehingga berdampak pada penurunan skor METEOR. Temuan ini menegaskan pentingnya evaluasi berbasis metrik linguistik seperti METEOR dibandingkan hanya mengandalkan indikator *loss* selama proses pelatihan.

Berdasarkan uji coba fungsional terbatas, sistem mampu menampilkan gambar dan *caption bilingual* secara konsisten, memproses unggahan gambar secara *real-time*, serta memutar audio dalam kedua bahasa tanpa kendala teknis. Dengan demikian, sistem yang dikembangkan tidak hanya berfungsi sebagai eksperimen akademik, tetapi juga sebagai *prototype* media pembelajaran digital yang mengintegrasikan modalitas visual, linguistik, dan auditori. Secara keseluruhan, seluruh tujuan penelitian yang telah dirumuskan dapat tercapai melalui proses perancangan, implementasi, dan evaluasi sistem. Sistem yang dihasilkan juga memberikan kontribusi dalam pengembangan *image captioning* bilingual berbasis *vision-language model* modern untuk mendukung pembelajaran pada tingkat pendidikan dasar.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya.

1. Penelitian selanjutnya dapat menggunakan ground truth yang dibuat secara manual oleh ahli bahasa atau pendidik untuk meningkatkan objektivitas evaluasi, sehingga tidak sepenuhnya bergantung pada pseudo-label dari model pretrained.
2. Evaluasi model dapat diperluas dengan menambahkan metrik lain seperti CIDEr, SPICE, atau BERTScore untuk memberikan analisis yang lebih komprehensif terhadap kualitas semantik caption.
3. Sistem dapat dikembangkan lebih lanjut dengan melakukan fine-tuning MarianMT atau menggunakan model multilingual yang lebih besar untuk meningkatkan kualitas translasi Bahasa Indonesia.
4. Penelitian lanjutan dapat melakukan uji eksperimen pendidikan secara terkontrol untuk mengukur dampak penggunaan sistem terhadap peningkatan literasi visual dan kemampuan bahasa siswa secara kuantitatif.
5. Sistem saat ini mendukung Bahasa Inggris dan Bahasa Indonesia, dan pada pengembangan selanjutnya dapat ditambahkan bahasa daerah maupun bahasa asing lainnya untuk memperluas cakupan pembelajaran bilingual atau multilingual.

Dengan pengembangan lebih lanjut, sistem image captioning bilingual berbasis vision–language model ini berpotensi menjadi media pembelajaran digital yang inovatif dan inklusif dalam mendukung literasi visual serta pembelajaran bilingual di sekolah dasar.

