

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Pengembangan generator soal HOTS berbasis *Small Language Model* berhasil diwujudkan melalui sebuah aplikasi web sederhana dengan fitur pilihan tingkat kelas, detail topik (misalnya angsa), jumlah soal dan level kesulitan. Pengguna dapat memilih parameter tersebut dan generator mengeluarkan beberapa pertanyaan yang menuntut keterampilan berpikir tingkat tinggi. Contoh soal yang dihasilkan seperti “Bagaimana kamu akan menganalisis peran kolam dalam pemeliharaan angsa dengan fokus pada keberlanjutan?”, yang menantang siswa menganalisis hubungan antar konsep dalam konteks lokal. Fitur-fitur ini menjawab rumusan masalah pertama: generator yang dirancang mampu menyesuaikan kebutuhan pembelajaran anak SD di Desa Pete dan menghasilkan soal kontekstual secara otomatis.

Untuk menjawab rumusan masalah kedua, kualitas soal dievaluasi menggunakan *BERTScore*, sebuah metrik kesamaan semantik. Hasil evaluasi awal menunjukkan bahwa *GPT-2* memiliki rata-rata F1-score 0,6382, lebih tinggi dari *ELECTRA-Small* (0,6143) dan *ALBERT-base-v2* (0,6285). Setelah proses fine-tuning, F1-score *GPT-2* meningkat signifikan menjadi sekitar 0,83 untuk setiap soal (misalnya 0,8344 pada soal kedua dan 0,8268 pada soal kelima). *ELECTRA-Small* meningkat menjadi rata-rata F1-score sekitar 0,75, dan *ALBERT-base-v2* menjadi sekitar 0,71. Hal ini menunjukkan adanya peningkatan akurasi dan relevansi soal setelah fine-tuning. Meski demikian, evaluasi melalui umpan balik guru belum dilakukan; langkah ini direncanakan saat aplikasi dipresentasikan di Desa Pete untuk memastikan soal yang dihasilkan benar-benar mengukur kemampuan analisis, evaluasi, dan kreasi siswa.

5.2 Saran

Berdasarkan kesimpulan, penelitian ini dapat dikembangkan lebih lanjut untuk meningkatkan kualitas sistem *Small Language Model* dalam pendidikan

berbasis kearifan lokal. Perluas dataset dengan transkrip video budidaya angsa yang lebih bervariasi, topik lokal seperti flora atau sejarah desa, serta teks dari buku pelajaran Ilmu Pengetahuan Alam (IPA) SD, agar soal HOTS lebih selaras dengan kurikulum. Tingkatkan evaluasi melalui uji coba langsung dengan siswa SD dan penilaian guru, di luar *BERTScore*, untuk mengukur dampak belajar. Bandingkan *GPT-2 Small* dengan model ringan lain seperti *DistilGPT-2*, *TinyLlama*, atau *Phi-3 Mini*, sambil optimalisasi fine-tuning untuk efisiensi di sumber daya terbatas. Kembangkan fitur aplikasi dengan manajemen pengguna, mode offline, dan teknologi mudah seperti *Laravel* atau *Laragon/XAMPP*, sehingga skalabel untuk desa-desa lain dan mendukung inovasi pendidikan AI inklusif.

5.3 Kontribusi Utama Penelitian

Penelitian ini memberikan tiga kontribusi penting. Pertama, dari sisi praktis, sistem generator soal HOTS berbasis SLM yang dikembangkan membantu guru di Desa Pete membuat soal yang relevan dengan konteks lokal budidaya angsa secara otomatis dan efisien, sehingga menghemat waktu dan memudahkan pengajar di desa lain dengan kondisi serupa. Kedua, secara metodologis, penelitian ini menerapkan kerangka CRISP-DM secara menyeluruh mulai dari pemahaman bisnis hingga *deployment* dalam pengolahan transkrip video lokal menjadi dataset, pemodelan, evaluasi, dan penerapan, memberikan contoh pipeline terstruktur yang dapat direplikasi untuk pengembangan *generator* HOTS serupa di konteks lain. Ketiga, kontribusi teoretisnya memperlihatkan bahwa model berukuran kecil seperti *GPT-2 Small*, ketika di-fine-tune pada data lokal, mampu menghasilkan soal HOTS yang efektif dengan peningkatan F1-score dari 0,6382 sebelum fine-tuning menjadi sekitar 0,83, sehingga menunjukkan bahwa *Small Language Models* dapat menjadi alternatif hemat sumber daya dibandingkan model besar untuk aplikasi pendidikan di daerah dengan keterbatasan teknologi.