BAB 5
KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Proses *fine-tuning* model Mistral-7B-Instruct berhasil dilakukan secara efisien pada *platform cloud* terbatas seperti Kaggle dengan menerapkan metode QLoRA. Hanya sekitar 92 juta parameter yang dilatih dari total 7,34 miliar, sehingga hemat sumber daya. Pelatihan dilakukan selama 1 *epoch* dengan 47 langkah optimisasi dan waktu komputasi sekitar 37 menit. Hasil pelatihan menunjukkan penurunan *training loss* dari 2,87 menjadi 0,99, serta *validation loss* akhir sebesar 0,94, yang menandakan proses pelatihan berjalan efektif meskipun menggunakan GPU terbatas.
2. Model hasil *fine-tuning* belum sepenuhnya optimal, namun telah menunjukkan kemampuan awal dalam menjawab pertanyaan berbasis dokumen akademik secara kontekstual. Evaluasi menggunakan BERTScore menghasilkan skor F1 sebesar 0.6421, dengan *Precision* 0.6323 dan Recall 0.6525. Meskipun akurasinya masih perlu ditingkatkan, model telah mampu memahami konteks pertanyaan dan menghasilkan jawaban yang relevan secara semantik.
3. Model berhasil dipublikasikan di Hugging Face Hub sehingga dapat digunakan ulang dan menjadi dasar pengembangan lebih lanjut. Hal ini membuka peluang untuk kolaborasi dan peningkatan model di masa depan, khususnya dalam konteks pengembangan sistem informasi akademik berbasis AI di lingkungan perguruan tinggi.
4. Implementasi prototipe *chatbot* belum berhasil diselesaikan secara penuh. Namun, model telah dipublikasikan dan siap digunakan sebagai *backend chatbot* pada tahap selanjutnya. Ini menunjukkan bahwa pondasi teknis

telah tersedia dan dapat dikembangkan lebih lanjut menjadi sistem chatbot akademik Unjaya.

5.2 SARAN

Berdasarkan keterbatasan dan temuan penelitian ini, beberapa saran untuk pengembangan lebih lanjut antara lain:

1. Perluasan dan diversifikasi dataset.

   Termasuk variasi gaya pertanyaan, kompleksitas jawaban, dan format respons agar model mampu menangani lebih banyak jenis input dan skenario nyata.

2. Eksperimen hyperparameter yang lebih komprehensif.

   Kombinasi seperti *learning rate*, *batch size*, dan parameter lain perlu diuji untuk mencapai performa terbaik terhadap karakteristik dataset Unjaya.

3. Evaluasi performa yang lebih luas.

   Selain BERTScore, disarankan menambahkan metrik lain seperti ROUGE, BLEU atau perplexity, serta evaluasi manual oleh manusia untuk mendapatkan gambaran kualitas yang lebih utuh. Perlu dilakukan riset lebih lanjut untuk mengevaluasi model secara menyeluruh dari berbagai aspek dan skenario penggunaan.

4. Pengembangan Sistem *Chatbot* yang Siap Digunakan

   Untuk mewujudkan sistem chatbot yang dapat digunakan secara nyata, perlu dilakukan pengembangan antarmuka pengguna berbasis web yang ramah pengguna, serta integrasi lebih lanjut dengan sistem informasi akademik yang relevan. Hal ini akan memperkuat nilai guna model dan mempermudah akses informasi bagi sivitas akademik.