

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem penjawab Tafsir Al-Qur'an berbasis *Hybrid Sparse-Dense Retrieval* dan *Retrieval-Augmented Generation (RAG)* dengan studi kasus *Tafsir Ibn Kathir*, dapat ditarik beberapa kesimpulan sebagai berikut.

1. Kinerja metode retrieval berbasis *sparse, dense, dan hybrid*

Hasil evaluasi menunjukkan bahwa metode *sparse retrieval (BM25)*, *dense retrieval (E5 Large Multilingual)*, dan pendekatan *hybrid sparse-dense* sama-sama mampu menghasilkan dokumen *Tafsir Ibn Kathir* yang relevan terhadap kueri pengguna. Meskipun nilai *MAP@10* bernilai maksimal pada seluruh metode, nilai *NDCG@10* menunjukkan bahwa pendekatan *hybrid* memberikan kualitas peringkat dokumen yang lebih baik, khususnya pada relevansi leksikal. Hal ini mengindikasikan bahwa penggabungan informasi leksikal dan semantik dapat meningkatkan kualitas hasil retrieval dibandingkan penggunaan satu pendekatan secara terpisah.

2. Performa sistem RAG dalam menghasilkan jawaban berbasis konteks tafsir

Sistem RAG yang dibangun mampu menghasilkan jawaban berdasarkan konteks dokumen *Tafsir Ibn Kathir* yang diperoleh dari tahap retrieval. Evaluasi menggunakan *semantic similarity*, *BERTScore*, dan *context relevance* menunjukkan bahwa jawaban yang dihasilkan memiliki kesesuaian semantik yang tinggi terhadap konteks, dengan tingkat konsistensi yang stabil pada konfigurasi *temperature* model *LLaMA 3.2:3B* sebesar 0.6. Pendekatan evaluasi berbasis embedding juga memastikan bahwa jawaban tetap berada dalam batas konteks dokumen tafsir yang digunakan.

3. Performa sistem dari sisi waktu respons (latensi)

Pengujian waktu respons menunjukkan bahwa sistem memiliki *end-to-end* latensi rata-rata sekitar 5.93 detik, dengan kontribusi terbesar berasal dari tahap generasi jawaban oleh model bahasa. Sementara itu, tahap retrieval memiliki waktu respons yang relatif lebih rendah. Hasil ini menunjukkan bahwa sistem dapat dijalankan sebagai prototipe aplikasi RAG berbasis Streamlit pada perangkat dengan spesifikasi terbatas, dengan waktu respons yang masih dapat diterima untuk skenario penggunaan interaktif.

4. Implementasi prototipe sebagai bukti konsep

Implementasi sistem dalam bentuk prototipe aplikasi *Streamlit* berhasil mengintegrasikan modul *retrieval*, modul generatif, dan antarmuka pengguna dalam satu alur kerja yang utuh. Prototipe ini dapat digunakan sebagai bukti konsep (*proof of concept*) serta sebagai dasar untuk evaluasi dan pengembangan sistem lebih lanjut di masa mendatang.

Secara keseluruhan, penelitian ini berhasil menjawab rumusan masalah yang diajukan dan mencapai tujuan penelitian, yaitu menganalisis kinerja metode *retrieval*, menilai performa sistem RAG berbasis tafsir, serta mengukur latensi sistem dalam implementasi prototipe.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemui selama proses pengerjaan, beberapa saran untuk pengembangan penelitian selanjutnya dapat disampaikan sebagai berikut.

1. Pengembangan metode evaluasi *retrieval*

Penelitian selanjutnya dapat menggunakan *ground truth* yang disusun oleh ahli tafsir atau evaluator independen untuk memperoleh penilaian relevansi yang lebih beragam dan mendalam, serta memungkinkan penggunaan metrik evaluasi tambahan seperti Recall@k atau MRR.

2. Eksplorasi model generatif dan strategi RAG lanjutan

Penggunaan model bahasa dengan ukuran parameter yang lebih besar atau pendekatan RAG lanjutan, seperti *reranking* berbasis LLM atau *multi-step retrieval*, dapat dieksplorasi untuk meningkatkan kualitas jawaban, terutama pada pertanyaan yang bersifat kompleks atau membutuhkan penalaran lintas konteks.

3. Optimasi performa dan latensi sistem

Untuk meningkatkan efisiensi waktu respons, penelitian selanjutnya dapat mempertimbangkan penggunaan *vector database*, mekanisme caching, atau pemrosesan paralel, terutama jika sistem diterapkan pada skala data dan jumlah pengguna yang lebih besar.

4. Perluasan cakupan data dan bahasa

Dataset tafsir dapat diperluas dengan menambahkan versi tafsir lain atau bahasa tambahan untuk meningkatkan cakupan pengetahuan sistem serta menguji kemampuan generalisasi pendekatan *hybrid retrieval* dan RAG pada konteks yang lebih luas.

Dengan adanya saran-saran tersebut, diharapkan penelitian ini dapat menjadi dasar bagi pengembangan sistem penjawab tafsir Al-Qur'an yang lebih komprehensif dan *robust* di masa mendatang.

