

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan evaluasi sistem tanya jawab pariwisata cerdas berbasis Retrieval-Augmented Generation (RAG) pada objek wisata Api Abadi Mrapen, Kabupaten Grobogan, dapat ditarik kesimpulan sebagai berikut:

1. **Keberhasilan Implementasi Sistem:** Sistem *Agentic* RAG berhasil dibangun menggunakan LangChain, LangGraph, FAISS, dan Groq API. Basis pengetahuan bersumber dari 643 ulasan pengunjung dan 20 data wawancara pengelola yang menghasilkan 738 vektor *embedding*. *Pipeline* beroperasi secara dinamis melalui 6 *node* dengan mekanisme *self-correction* (ambang batas *cosine similarity* 0,30).
2. **Presisi Pencarian Cukup Baik (*Cosine Similarity*):** Evaluasi kedekatan semantik mencatatkan rata-rata skor 0,7039 (Relevan). Dari 8 pengujian, 37,5% berskor Sangat Relevan, 37,5% Relevan, dan 25% Kurang Relevan—khususnya pada kueri eksploratif terkait area parkir (0,544) dan objek wisata sekitar (0,484).
3. **Variasi Kinerja pada Evaluasi RAGAS Dasar:** Pengujian awal dengan 10 *test case* menghasilkan skor global 0,656. Sistem berhasil melampaui target pada metrik *Faithfulness* (0,916) dan *Context Recall* (0,750), meski tiga metrik lainnya (*Answer Relevancy*, *Context Precision*, dan *Answer Correctness*) belum mencapai nilai ideal. Kinerja paling optimal tercatat pada kueri penjual makanan (TC-05; rata-rata 0,837).
4. **Keunggulan *Agentic* vs *Baseline* RAG:** Pada skala 30 *test case* gabungan, *Agentic* RAG unggul dengan skor global 0,597 dibanding *Baseline* (0,576), atau surplus 3,65%. Sesuai hipotesis, logika *self-correction* memberikan nilai tambah konsisten pada *Context Precision*, *Context Recall*, dan *Answer Correctness*. Namun, keunggulan tipis *Baseline* pada metrik *Faithfulness* (0,793 berbanding 0,779) menandakan *pipeline* *Agentic* masih memerlukan kalibrasi lanjutan untuk menjaga kesetiaan jawaban terhadap konteks.
5. **Penerimaan Pengguna Sangat Tinggi (UAT):** Pengujian kualitatif kepada 34 responden lintas profil meraih skor 4,42/5,00 (Sangat Baik). Aspek Kepuasan Umum

memimpin dengan skor tertinggi (4,50). Secara praktis, 91,2% pengguna merasakan manfaat langsung sistem ini dan 67,6% bersedia merekomendasikannya kepada orang lain.

5.2 Saran

Berdasarkan temuan dan keterbatasan yang diidentifikasi selama penelitian, berikut saran untuk pengembangan sistem dan penelitian selanjutnya:

1. **Perluasan Kata Kunci Domain:** Menambah variasi kata (seperti "padam", "metana", "PVMBG") pada *node cek_relevansi* untuk meminimalisasi kesalahan klasifikasi (*fallback*) pada pertanyaan spesifik (seperti kasus TC-15).
2. **Optimasi *Chunking* dan *Retrieval*:** Menerapkan strategi *semantic* atau *hierarchical chunking*, serta penyesuaian nilai *k* pada *retriever* secara dinamis guna memaksimalkan capaian *Context Precision*.
3. **Penyempurnaan Logika *Self-Correction*:** Mengkaji ulang *threshold cosine similarity* dan logika *grading* pada LangGraph agar metrik *Faithfulness* Agentic RAG dapat meningkat tanpa mengorbankan keunggulannya pada presisi dan kelengkapan. Metrik *Answer Relevancy* juga perlu diikutsertakan pada pengujian skala besar.
4. **Perluasan Pengujian Pengguna (UAT):** Melaksanakan UAT langsung di lokasi (Api Abadi Mrapen) dengan melibatkan wisatawan, pemandu, dan pengelola. Penggunaan instrumen terstandarisasi seperti *System Usability Scale* (SUS) atau *Net Promoter Score* (NPS) sangat disarankan.
5. **Pembaruan Basis Pengetahuan Dinamis:** Mengimplementasikan *incremental update* pada *vector store* FAISS untuk efisiensi penambahan data baru, serta memperkaya referensi dengan rilis resmi dari Pemkab Grobogan atau laporan PVMBG.
6. **Replikasi Lintas Objek Wisata:** Mengembangkan kerangka kerja arsitektur *Agentic RAG* menjadi *framework* generik yang mudah dikonfigurasi ulang untuk objek wisata alam lainnya di kawasan Kabupaten Grobogan maupun Jawa Tengah.