

MODEL CHATBOT INFORMASI AKADEMIK ITK BERBASIS RAG MENGGUNAKAN LIGHTWEIGHT LLAMA

Nama Mahasiswa : Muhammad Saman
NIM : 11211052
Dosen Pembimbing Utama : Gusti Ahmad Fanshuri Alfarisy, S.Kom.,
M.Kom.
Dosen Pembimbing Pendamping : Rizky Amelia, S.Si., M.Han.

ABSTRAK

Dalam era digital yang serba cepat, Institut Teknologi Kalimantan (ITK) menghadapi tantangan dalam menyediakan informasi akademik yang cepat, akurat, dan mudah diakses oleh mahasiswa, dosen, serta staf akademik. Untuk mengatasi permasalahan keterbatasan akses terhadap informasi seperti prosedur administrasi, tata cara penulisan laporan, dan kebijakan akademik yang masih bergantung pada layanan manual atau handbook, penelitian ini mengembangkan chatbot berbasis *Retrieval-Augmented Generation* (RAG) dengan model LLaMA. Chatbot ini menggabungkan kemampuan pencarian informasi dan pemrosesan bahasa alami guna memberikan jawaban yang relevan dan kontekstual. Berdasarkan evaluasi menggunakan metrik RAGAS, *Mean Reciprocal Rank* (MRR), dan latensi, konfigurasi terbaik diperoleh pada kombinasi model LLaMA 3B dengan Indo-Sentence-BERT, ukuran *chunk* 200 token dan *overlap* 10 token, dengan skor RAGAS tertinggi sekitar 0,9, MRR sekitar 0,5, serta latensi stabil di bawah 1 detik. Meskipun LLaMA 1B menunjukkan MRR lebih tinggi (0,6), performa RAGAS-nya rendah sehingga tidak direkomendasikan. Dengan mempertimbangkan seluruh metrik, konfigurasi LLaMA 3B dengan Indo-Sentence-BERT dipilih sebagai solusi utama. Chatbot ini diharapkan mampu meningkatkan efisiensi pencarian informasi akademik di ITK secara signifikan.

Kata kunci: *Retrieval-Augmented Generation* (RAG), LLaMA, Institut Teknologi Kalimantan (ITK), RAGAS, MRR, Latensi