

**CHATBOT UNTUK MENDUKUNG PENCARIAN INFORMASI
PADA DOKUMEN JURNAL ILMIAH INFORMATIKA DENGAN RAG
DAN MODEL *LIGHTWEIGHT* LLaMA 3.2**

Nama Mahasiswa : Dwi Savira Aprilianti
NIM : 11211033
Dosen Pembimbing Utama : Gusti Ahmad Fanshuri Alfarisy, Ph.D.
Dosen Pembimbing Pendamping : Ramadhan Paninggalih S.Si., M.Si., M.Sc

ABSTRAK

Meningkatnya jumlah dokumen jurnal ilmiah informatika menimbulkan tantangan dalam memperoleh informasi yang cepat, akurat, dan sesuai konteks. Penelitian ini bertujuan mengembangkan dan mengevaluasi sistem *chatbot* berbasis *Retrieval-Augmented Generation* (RAG) dengan model LLaMA 3.2 berparameter 1B dan 3B untuk mendukung pencarian informasi pada dokumen jurnal ilmiah informatika, serta menentukan nilai *threshold cosine similarity* terbaik berdasarkan *Accuracy* dalam mendeteksi ketidaktahuan. Dataset penelitian berupa tiga dokumen jurnal ilmiah informatika berformat JATS/XML. Proses pengembangan meliputi *parsing*, *preprocessing*, *chunking*, pembentukan *embedding* menggunakan *all-MiniLM-L6-v2*, penyimpanan pada *ChromaDB*, *retrieval*, *reranking*, dan pembangkitan jawaban menggunakan LLaMA 3.2. Evaluasi dilakukan menggunakan *Mean Reciprocal Rank* (MRR), RAGAS, serta pengujian *threshold cosine similarity*. Hasil penelitian menunjukkan bahwa proses pengelolaan dokumen menghasilkan 257 dokumen dan 269 *chunk*. Evaluasi *retrieval* terhadap 30 *query* dengan $\text{top-k} = 5$ memperoleh nilai MRR sebesar 0,8667. Model LLaMA 3.2 1B memperoleh nilai *Faithfulness* 0,9611, *Answer Relevancy* 0,7655, dan *Context Relevance* 0,9250, sedangkan model 3B memperoleh *Faithfulness* 0,9185, *Answer Relevancy* 0,8098, dan *Context Relevance* 0,9250. Pengujian menunjukkan *threshold cosine similarity* sebesar 0,40 memberikan hasil terbaik dengan *Accuracy* 0,9778 dan *Balanced Accuracy* 0,9833. Hasil tersebut menunjukkan bahwa *chatbot* berbasis RAG dan LLaMA 3.2 mampu mendukung pencarian informasi secara efektif sekaligus mendeteksi pertanyaan di luar cakupan pengetahuan dengan baik.

Kata kunci: *Chatbot*, *Retrieval-Augmented Generation*, LLaMA 3.2, *threshold cosine similarity*, deteksi ketidaktahuan.