ABSTRAK

Marcelino Gavriel:

Skripsi

Implementasi Sistem API Chatbot untuk customer PT XYZ dengan Large Language Model menggunakan metode Hybrid Search RAG

PT XYZ adalah perusahaan agri-food di Indonesia yang memproduksi pakan ternak, bibit unggul, peternakan ayam, udang, ikan, hingga olahan hewan. Proses layanan informasi kepada customer masih dilakukan secara manual oleh admin customer service, sehingga sering terjadi keterlambatan respons dan pertanyaan yang berulan. Oleh karena itu, dalam tugas akhir kali ini dikembangkan sistem API chatbot sebagai solusi awal untuk membantu PT XYZ dalam menangani pertanyaan customer secara otomatis. Sistem ini juga dilengkapi web base admin untuk menambah dan memperbarui informasi pada database chatbot. Chatbot dibangun menggunakan model Llama-3.1 8B ArliAI IndoFormax dengan metode hybrid search Retrieval Augmented Generation (RAG), yang menggabungkan similarity search dan keyword search menggunakan pgvector, serta reranking sebelum konteks masuk ke dalam pipeline LLM.

Web base admin dikembangkan dengan framework Streamlit dan basis data PostgreSQL, serta menggunakan Sentence Transformer Indo-Sentence-BERT untuk embedding informasi. Pengujian sistem dilakukan bersama tim admin dan IT PT XYZ untuk mengevaluasi kemampuan chatbot dalam memberikan jawaban yang relevan serta kemudahan pengelolaan data melalui web base yang telah dibuat.

Kata kunci : chatbot, api, llm, hybrid search, similarity search, keyword search, reranking

ABSTRACT

Marcelino Gavriel:

Undergraduate thesis

Implementation of a Chatbot API System for PT XYZ Customers Using a Large Language Model with the Hybrid Search RAG Method

PT XYZ isf an agri-food company in Indonesia producing animal feed, superior livestock, poultry, shrimp, fish, and processed animal products. Customer inquiries regarding products and sales information are still handled manually by customer service staff, often causing delayed responses and repeated questions. Therefore, in this final project, an API-based chatbot system was developed as an initial solution to help PT XYZ manage customer inquiries automatically. Additionally, a web-based admin system was provided to enable administrators to add and update information in the chatbot's database. The chatbot is built using the Llama-3.1 8B ArliAI IndoFormax model with a hybrid search Retrieval Augmented Generation (RAG) method, combining similarity search and keyword search using pgvector, as well as reranking before the context is processed through the LLM pipeline to generate relevant answers.

The admin web system was developed using the Streamlit framework and PostgreSQL database, utilizing the Indo-Sentence-BERT Sentence Transformer for information embedding. System testing was conducted together with PT XYZ's admin and IT teams to evaluate the chatbot's ability to provide accurate responses and the ease of managing chatbot data through the web-based system.

keyword sentences : chatbot, api, llm, hybrid search, similarity search, keyword search, reranking

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL