Food Nutrition Detection

Problem Definition

🎯 Goals

Business Goals

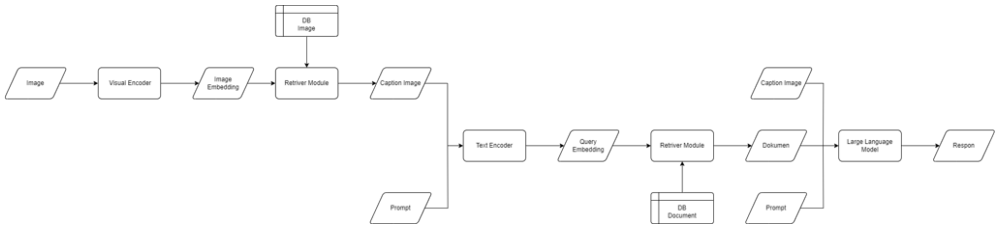
How the system can detect food through images and provide nutritional content information, and is able to provide relevant interactive responses regarding food.

System Goals

- This system is able to recognize and classify more than one type of Indonesian food in one image.
- The system is able to provide nutritional content information based on food images.
- The system is able to provide responses in a natural and easy-to-understand language style.
- The system is capable of providing multiple responses to a single image through multi-turn conversational interactions.
- The system is capable of running reasoning and question and answer (Q&A) processes related to food and nutrition information using integration with the Large Language Model (LLM).
- The system is capable of generating responses in Indonesian.
- The system is able to run without requiring an external server.

📋 Planner

Pipeline AI (LLaMA + RAG)



Explanation:

- Text Encoder: Converting text to vector embedding
- Visual Encoder: Converting image to vector embedding
- Retriver Module: take captions that suit the image and documents that are relevant to the text
- Large languange Model: Process text prompts and produce output in the form of new text based on knowledge, instructions, or questions.

🏁 Execution

Model

- Text Encoder: Sentence-BERT
- Visual Encoder : Dinov2
- Large Language Model : LLaMA 3

Database

PostgresSQL

Framework

Pytorch, LangChain