BAB I PENDAHULUAN

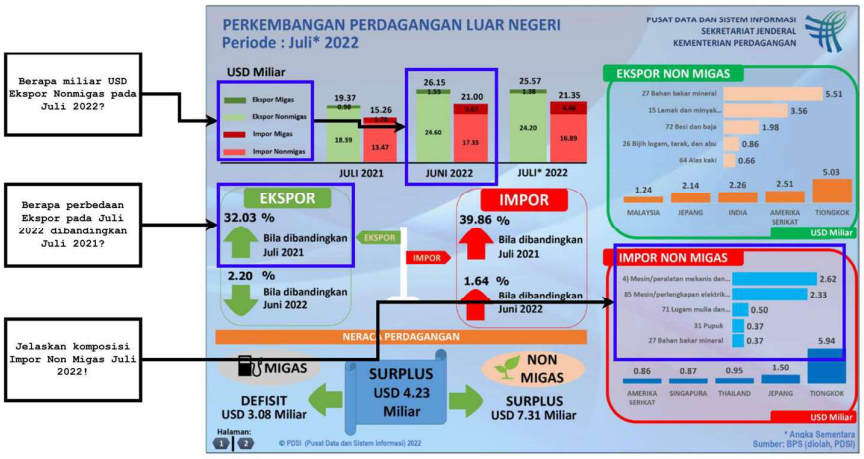
I.1 Latar Belakang

Digitalisasi yang terus berkembang pesat di berbagai sektor, khususnya di sektor keuangan, telah mendorong perusahaan untuk mengadopsi teknologi guna meningkatkan efisiensi dan akurasi pengelolaan data (Yang, 2022). Saat ini, BRI memiliki lebih dari 70 juta nasabah dengan lebih dari 40 juta pengguna BRImo. Jumlah ini terus bertambah setiap tahun (Anam, 2025). Hal ini menyebabkan masalah yang dihadapi organisasi tersebut menjadi makin beragam, mulai dari kendala teknis aplikasi, pertanyaan seputar transaksi, hingga isu verifikasi data. Setiap kelompok nasabah punya karakteristik dan kebutuhan layanan yang berbeda. Untuk menangani permasalahan tersebut secara efisien, staf BRI perlu dibekali bahan ajar yang sesuai konteks.

Dalam praktiknya, bahan ajar disusun dalam berbagai format, seperti materi presentasi, infografis, panduan teknis, dan simulasi percakapan. Keragaman ini sejalan dengan kompleksitas permasalahan yang dihadapi nasabah, serta perbedaan kebutuhan berdasarkan jalur layanan dan segmentasi pengguna. Untuk memastikan bahan ajar dapat digunakan secara efisien dalam konteks operasional, diperlukan sistem pencarian informasi yang mampu mengakses dan menginterpretasi informasi lintas format secara akurat dan cepat. Teknologi tradisional seperti *search engine* cenderung tidak efisien, karena staf tetap harus membaca dan menyaring informasi secara manual sebelum menemukan jawaban yang relevan (Lewis dkk., 2020). Sementara itu, pendekatan RAG berbasis teks saja sering kali menghasilkan respons yang kurang akurat akibat hilangnya konteks visual, yang dalam banyak kasus operasional justru memuat informasi kunci (Cho dkk., 2024; Mathew, Karatzas, dkk., 2021).

Untuk menjawab tantangan tersebut, diperlukan penerapan teknologi berbasis kecerdasan buatan (AI) yang lebih maju. Salah satu pendekatan terbaru adalah *Multimodal Retrieval Augmented Generation* (MRAG), yang memanfaatkan model seperti ColPali dan *Vision Language Model*. Teknologi ini dirancang untuk mengekstraksi, memahami, dan mengintegrasikan informasi dari berbagai jenis dokumen dengan lebih efektif. ColPali unggul dalam memproses teks dan data

terstruktur dalam task *document retrieval* (Faysse dkk., 2024), sementara *Vision Language Model* memiliki kemampuan *multimodal* yang memungkinkan analisis data dalam berbagai format, termasuk gambar dan dokumen dengan tata letak kompleks (Lin dkk., 2023; Xia dkk., 2024). Kombinasi kedua teknologi ini memberikan potensi besar dalam pengembangan sistem yang mampu mengubah data tidak terstruktur menjadi informasi bernilai tinggi.



Gambar I-1 Contoh Dokumen Visual

Gambar I-1 mengilustrasikan contoh dokumen visual kompleks berupa infografik mengenai perkembangan perdagangan luar negeri Indonesia. Gambar ini menunjukkan bagaimana informasi tekstual dan grafis (seperti diagram batang) disajikan secara terintegrasi. Pemrosesan dokumen semacam ini tidak efektif jika hanya mengandalkan pendekatan berbasis teks, karena makna dan konteksnya sangat bergantung pada tata letak visual, hubungan antar-elemen grafis, dan teks yang menyertainya. Contoh tersebut menunjukkan pentingnya sistem *multimodal* yang mampu memahami dan mengintegrasikan informasi dari kedua modalitas tersebut secara simultan untuk menjawab pertanyaan spesifik secara akurat, seperti yang dicontohkan oleh contoh pertanyaan pada gambar.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengidentifikasi dan mengembangkan sistem *Multimodal Retrieval Augmented Generation* berbasis ColPali dan *Vision Language Model*. Dengan pendekatan ini,

diharapkan bahwa sistem tersebut mampu mengatasi hambatan pengolahan data tidak terstruktur berupa kumpulan file PDF sehingga dapat memberikan jawab berbasis fakta dari *knowledge* internal dengan kualitas yang lebih tinggi dari sistem berbasis teks untuk meningkatkan efisiensi operasional (Faysse dkk., 2024; Lewis dkk., 2020; Lin dkk., 2023).

I.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, PT Bank Rakyat Indonesia (BRI) menghadapi permasalahan dalam pengolahan data yang bersumber dari dokumen-dokumen yang tidak terstruktur, seperti gambar, pdf *file*, dokumen presentasi, serta *scan* dokumen fisik. Ketidakakuratan dalam pemrosesan data yang disebabkan oleh keterbatasan sistem dalam memahami konten dokumen yang kompleks, seperti teks tidak terstruktur, tabel, dan gambar, dapat mengakibatkan kesalahan dalam ekstraksi informasi penting. Selain itu, inefisiensi juga terjadi karena metode RAG berbasis teks masih bergantung pada serangkaian proses panjang seperti *layout detection*, OCR, dan *chunking* untuk memecah dokumen, yang tidak hanya memakan waktu tetapi juga meningkatkan kompleksitas komputasi. Kedua kendala ini dapat menghambat efektivitas proses bisnis, memperlambat alur kerja, dan menurunkan kualitas hasil analisis. Untuk itu, diperlukan solusi yang efektif berupa sistem *Multimodal Retrieval Augmented Generation* dengan ColPali dan *Vision Language Model*.

Rumusan masalah yang mendasari penelitian ini adalah:

1. Bagaimana cara merancang dan mengembangkan model ColPali yang mampu menghasilkan embedding dokumen berbasis representasi *dense multivector* untuk mendukung pencarian informasi pada dokumen teks tidak terstruktur berupa kumpulan file pdf di PT Bank Rakyat Indonesia?
2. Bagaimana cara mengembangkan *Vision Language Model* (VLM) yang dapat memproses dokumen *multimodal* secara *end-to-end* tanpa tahap pra-pemrosesan seperti OCR dan *layout parsing*, sehingga mampu memahami informasi visual dan tekstual secara bersamaan?

3. Bagaimana cara membangun sistem *Multimodal Retrieval Augmented Generation* (MRAG) yang mengintegrasikan ColPali dan VLM untuk meningkatkan pemanfaatan *knowledge* internal, mendukung proses analisis, dan membantu pengambilan keputusan berbasis dokumen *multimodal* di lingkungan PT Bank Rakyat Indonesia?

I.3 Tujuan Tugas Akhir

Penelitian ini bertujuan untuk:

1. Memahami data, mengolah data, dan merancang model dengan melakukan *fine-tuning*, dan mengevaluasi model ColPali untuk menghasilkan *embedding* dokumen berbasis representasi *dense multivector* yang dioptimalkan untuk pencarian informasi pada dokumen teks tidak terstruktur dalam konteks internal perusahaan.
2. Memahami data, mengolah data, dan merancang model dengan melakukan *fine-tuning*, dan mengevaluasi *Vision Language Model* (VLM) agar mampu memahami dokumen multimodal tanpa tahap pra-pemrosesan seperti OCR dan *layout parsing*, sehingga dapat memproses informasi visual dan tekstual secara bersamaan.
3. Merencanakan, membangun, mengintegrasikan, dan menguji sistem *Multimodal Retrieval-Augmented Generation* (MRAG) yang menggabungkan ColPali dan VLM untuk meningkatkan efisiensi pemanfaatan *knowledge* internal perusahaan, mendukung analisis data, serta membantu pengambilan keputusan berbasis dokumen *multimodal* di PT Bank Rakyat Indonesia.

I.4 Manfaat Tugas Akhir

Penelitian ini diharapkan dapat memberikan manfaat signifikan baik secara teoretis bagi komunitas akademis maupun secara praktis bagi industri perbankan, yang dijabarkan sebagai berikut:

1. Bagi Komunitas Riset Akademis, penelitian ini menghasilkan sebuah *dataset image-question-answer* baru dalam Bahasa Indonesia yang secara spesifik dibangun dari dokumen-dokumen visual yang kompleks. *Dataset* ini dapat menjadi sumber daya dan *benchmark* yang berharga untuk mendorong penelitian selanjutnya dalam bidang NLP dan *computer vision*, khususnya untuk bahasa dengan sumber daya terbatas seperti Bahasa Indonesia.

2. Bagi *Open Source LLM Community*, penelitian ini menyumbangkan artefak model yang telah diadaptasi secara efektif untuk Bahasa Indonesia. Secara spesifik, penelitian ini menyediakan versi *fine-tuned* dari model *retriever multimodal* (ColQwen2.5) dan *Vision Language Model* (Qwen2.5-VL),
3. Bagi PT Bank Rakyat Indonesia (BRI) dan Industri Serupa, manfaat utama adalah pengembangan sistem *Multimodal Retrieval Augmented Generation* (MRAG) fungsional yang mampu mengoptimalkan manajemen pengetahuan internal. Secara praktis, sistem ini dapat:

a. Meningkatkan efisiensi operasional dengan memangkas waktu yang dibutuhkan karyawan untuk mencari dan memahami informasi dari dokumen tidak terstruktur berupa kumpulan file PDF secara manual.
b. Mendukung proses pengambilan keputusan yang lebih akurat dengan meminimalkan risiko kesalahan interpretasi data dari dokumen visual yang kompleks.
c. Menyediakan cetak biru arsitektur berbasis *microservices* dan hybrid *cloud* yang andal dan skalabel untuk implementasi teknologi AI serupa di lingkungan *enterprise*.

I.5 Batasan dan Asumsi Tugas Akhir

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan dalam pengembangan dan implementasi sistem *Multimodal Retrieval Augmented Generation* (MRAG), yaitu:

1. Sistem RAG yang dikembangkan hanya difokuskan pada pengolahan data teks dan gambar, dan belum mencakup jenis media lain seperti audio atau video, yang mungkin diperlukan dalam konteks komunikasi *multimodal* yang lebih kompleks.

2. Sistem ini hanya dioptimalkan untuk menjawab pertanyaan dan memberikan informasi berdasarkan dokumen yang telah tersedia dalam format teks dan gambar. Oleh karena itu, interaksi dengan data atau sumber informasi lain yang lebih dinamis, seperti video atau data *real time*, tidak menjadi fokus utama dalam penelitian ini.

Batasan-batasan ini dimaksudkan untuk menjaga penelitian tetap terarah dan fokus pada kebutuhan PT Bank Rakyat Indonesia dalam meningkatkan efisiensi pengolahan dokumen yang paling relevan.

I.6 Sistematika Laporan

Laporan tugas akhir ini disusun secara sistematis untuk memberikan pemahaman yang komprehensif mengenai pengembangan sistem *Multimodal Retrieval Augmented Generation* (MRAG) dengan ColPali dan *Vision Language Model* pada PT Bank Rakyat Indonesia. Sistematika penulisan laporan ini adalah sebagai berikut:

BAB I: PENDAHULUAN

Bab ini menyajikan gambaran umum penelitian, dimulai dengan latar belakang yang menguraikan tantangan pengelolaan dokumen tidak terstruktur di PT Bank Rakyat Indonesia dan potensi solusi menggunakan teknologi AI. Selanjutnya, dirumuskan permasalahan spesifik yang dihadapi, diikuti dengan penetapan tujuan penelitian yang ingin dicapai. Bab ini juga menjelaskan batasan-batasan penelitian yang ditetapkan untuk menjaga fokus studi, serta manfaat yang diharapkan dari hasil penelitian bagi PT Bank Rakyat Indonesia, industri perbankan, dan peneliti lain.

BAB II: TINJAUAN PUSTAKA

Bab ini membahas landasan teori yang relevan dan penelitian terdahulu yang menjadi dasar pengembangan sistem. Uraian mencakup konsep-konsep inti seperti arsitektur *Transformer*, *Large Language Models* (LLM), *Low-Alpha Adaptation* (LoRA), Information Retrieval (IR), *Retrieval Augmented Generation* (RAG), *Vision Language Model*s (VLM), dan secara khusus membahas *Multimodal Retrieval Augmented Generation* menggunakan ColPali. Selain itu, dibahas pula arsitektur pendukung seperti RESTful API dan *Microservice Architecture*. Bagian akhir bab ini menyajikan justifikasi pemilihan teori, kerangka kerja, dan mekanisme yang digunakan dalam penelitian ini.

BAB III: METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci pendekatan dan langkah-langkah metodologis yang digunakan dalam pelaksanaan penelitian. Pada bab ini dijelaskan metodologi CRISP-DM untuk aspek *data mining*. Bab ini juga merinci metode pengumpulan data, tahapan pengolahan data dan pengembangan produk yang meliputi Data ETL, modeling (termasuk *fine-tuning* ColPali dan VLM), desain sistem, dan pengembangan sistem menggunakan teknologi *cloud*. Terakhir, dijelaskan metode evaluasi yang akan digunakan untuk mengukur keberhasilan sistem.

BAB IV: ANALISIS DAN PERANCANGAN

Bab ini berfokus pada proses analisis kebutuhan dan perancangan sistem *Multimodal* RAG. Diawali dengan analisis proses bisnis sebelum dan sesudah implementasi sistem yang diusulkan untuk menunjukkan potensi peningkatan efisiensi. Selanjutnya, dipaparkan proses pengumpulan data spesifik untuk penelitian ini, termasuk sumber data dan pembuatan label sintetik menggunakan GPT-4o.

Bagian inti dari bab ini adalah *detail* modeling, yang mencakup proses *fine-tuning* model *retriever* ColQwen2.5 dan *fine-tuning Vision Language Model* Qwen2.5-VL. Kemudian, disajikan perancangan sistem secara menyeluruh, meliputi desain arsitektur *microservices* untuk *Embedding Service*, *Vector Database* (Qdrant), *Indexing Scheduler*, *Document Retriever Endpoint*, LLM *Service*, RAG *Endpoint*, pipeline end-to-end, dan arsitektur sistem secara keseluruhan yang mengintegrasikan infrastruktur *on-premise* dengan *cloud*.

BAB V: EVALUASI SISTEM

Bab ini menyajikan hasil pengujian dan evaluasi terhadap sistem yang telah dikembangkan. Evaluasi dilakukan secara terpisah untuk komponen *document retriever*, di mana kinerja ColQwen2.5 hasil *fine-tuning* dibandingkan dengan model baseline lainnya menggunakan metrik seperti Recall@k dan MRR@5.

Selanjutnya, dilakukan evaluasi terhadap kinerja generasi jawaban oleh *Vision Language Model* (Qwen2.5-VL hasil *fine-tuning*) menggunakan metrik BERTScore dan akurasi berdasarkan LLM-Eval, juga dengan perbandingan

terhadap model lain. Bab ini juga menyertakan pembahasan mengenai evaluasi integrasi sistem secara keseluruhan dalam lingkungan produksi, meskipun *detail* UAT bersifat internal.

BAB VI: KESIMPULAN DAN SARAN

Bab terakhir ini merangkum keseluruhan hasil penelitian dan menarik kesimpulan berdasarkan analisis yang telah dilakukan. Kesimpulan menjawab rumusan masalah dan tujuan penelitian yang telah ditetapkan di Bab I, serta menyoroti kontribusi utama penelitian, yaitu *dataset image-question*-answer baru untuk bahasa Indonesia dan peningkatan kinerja sistem *multimodal* OpenQA melalui adaptasi model. Selain itu, bab ini juga memberikan saran-saran konstruktif untuk pengembangan lebih lanjut dan penelitian di masa mendatang guna meningkatkan kapabilitas dan cakupan sistem.