

ABSTRAK

PT Bank Rakyat Indonesia (BRI) menghadapi tantangan dalam mengolah dan mengakses informasi dari volume besar dokumen *multimodal* tidak terstruktur yang mengandung teks dan gambar. Sistem manual atau pendekatan *Retrieval Augmented Generation* (RAG) tradisional seringkali tidak efisien dan dapat kehilangan konteks visual penting. Penelitian ini bertujuan untuk mengembangkan sistem *Multimodal RAG* (MRAG) guna meningkatkan akurasi dan efisiensi ekstraksi informasi dari dokumen-dokumen tersebut di BRI. Pada penelitian ini digunakan *framework Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang meliputi pemahaman bisnis dan data, persiapan data, pemodelan, evaluasi, dan *deployment* sistem. Sistem yang diusulkan mengadaptasi pendekatan ColPali, diimplementasikan sebagai ColQwen2.5, untuk *document retrieval* dengan memperlakukan setiap halaman dokumen sebagai gambar. Selain itu, sistem ini memanfaatkan *Vision Language Model* (VLM) Qwen2.5-VL yang telah di-*fine-tuning* untuk generasi jawaban yang relevan secara kontekstual. Kontribusi utama dari penelitian ini adalah pembuatan *dataset image-question-answer* baru dalam bahasa Indonesia, yang dikumpulkan dari dokumen publik visual dan terdiri dari 1.318 gambar tunggal serta 3.930 pasangan tanya-jawab. Hasil evaluasi menunjukkan performa yang signifikan: model *retriever* ColQwen2.5 mencapai $MRR@5$ sebesar 0,92762, sementara model generator Qwen2.5-VL hasil *fine-tuning* mencapai skor BERT-F1 0,8534 dan akurasi LLM-Eval (menggunakan GPT-4o) sebesar 0,8603, menunjukkan peningkatan 3,3% dibandingkan model dasarnya. Pengembangan sistem MRAG ini menawarkan potensi besar untuk mengoptimalkan manajemen pengetahuan dan mendukung proses pengambilan keputusan yang lebih baik di PT Bank Rakyat Indonesia.

Kata kunci— *Multimodal Retrieval Augmented Generation, ColPali, Vision Language Model, Information Retrieval, Visual Document Understanding.*