

ABSTRAK

Jumlah dokumen pengetahuan perkeretaapian di PT Kereta Api Indonesia (Persero) terus meningkat setiap tahunnya. Namun, peningkatan tersebut belum diimbangi dengan ketersediaan sistem pengelolaan pengetahuan yang mampu memahami dokumen secara kontekstual serta analisis evaluasinya. Atas dasar hal tersebut, penelitian ini bertujuan untuk mengembangkan *chatbot Knowledge Management System* berbasis *Retrieval-Augmented Generation* dengan *framework* RAGOps sebagai tata kelola operasional, serta mengintegrasikan mekanisme *auto evaluation* dengan RAG Triad untuk memastikan kualitas jawaban sistem secara otomatis. Dataset yang digunakan merupakan korpus perkeretaapian multistruktur yang bersifat publik, terdiri atas pasangan tanya–jawab serta dokumen sumber. Hasil pengujian menunjukkan bahwa RAGAS merupakan *library auto evaluation* paling relevan karena memiliki korelasi Pearson tertinggi terhadap penilaian kebenaran ($r = 0,66$ untuk *Naive* dan $r = 0,69$ untuk *Advance*), diikuti Azure AI Evaluation SDK ($r = 0,62$), sementara DeepEval memiliki korelasi rendah ($r = 0,24$ dan $r = 0,17$). Kalibrasi ambang batas menggunakan Indeks Youden menghasilkan *cutoff* optimal 0,72 untuk *Naive* dan 0,6 untuk *Advance* yang menjadi dasar pengambilan keputusan otomatis dalam proses evaluasi. Evaluasi performa menunjukkan bahwa konfigurasi *Hybrid* memberikan hasil terbaik dengan akurasi 0,86, *precision* 0,85, *recall* 1, dan F1-score 0,92, lebih tinggi dibanding F1-score *Naive* sebesar 0,84 dan *Advance* sebesar 0,85. Dari sisi biaya, penerapan pola dari *Naive* ke *Hybrid* menunjukkan nilai *Incremental Benefit–Cost Ratio* (IBCR) sebesar 29,6 dan *Incremental Cost–Effectiveness Ratio* (ICER) sebesar \$0,034 per +1 F1-score atau \$0,00034 per +0,01 F1-score. Peningkatan biaya yang dikeluarkan selaras dengan peningkatan performa sistem. Penerapan mekanisme *auto evaluation* berbasis metrik RAG Triad pada *framework* RAGOps dan mampu meningkatkan performa *chatbot* KMS perkeretaapian.

Kata kunci—Sistem Manajemen Pengetahuan, RAGOps, *LLM-as-a-Judge*, RAG Triad, Perkeretaapian

ABSTRACT

The number of railway knowledge documents at PT Kereta Api Indonesia (Persero) continues to increase every year. However, this increase has not been matched by the availability of a knowledge management system capable of understanding documents contextually and evaluating them. Based on this, this study aims to develop a Retrieval-Augmented Generation-based Knowledge Management System chatbot with the RAGOps framework as operational management, and to integrate an auto-evaluation mechanism with RAG Triad to ensure the quality of the system's responses automatically. The dataset used is a public, multi-structured railway corpus consisting of question–answer pairs and source documents. The experimental results show that Ragas is the most relevant auto-evaluation library, demonstrated by the highest Pearson correlation with correctness assessment ($r = 0,66$ for the Naive configuration and $r = 0,69$ for the Advance configuration), followed by the Azure AI Evaluation SDK ($r = 0,62$), while DeepEval exhibits low correlation ($r = 0,24$ and $r = 0,17$). Threshold calibration using the Youden Index yields optimal cutoffs of 0,72 for the Naive pipeline and 0,6 for the Advance pipeline, which serve as the basis for automated decision making in the evaluation process. Performance evaluation shows that the Hybrid configuration provides the best results with an accuracy of 0,86, precision of 0,85, recall of 1, and F1-score of 0,92, higher than the Naive F1-score of 0,84 and Advance F1-score of 0,85. From a cost perspective, the transition from Naive to Hybrid demonstrates an Incremental Benefit–Cost Ratio (IBCR) of 29,6 and an Incremental Cost–Effectiveness Ratio (ICER) of \$0,034 per +1 F1-score or \$0,00034 per +0,01 F1-score. The increase in cost is proportional to the improvement in system performance. The implementation of an auto-evaluation mechanism based on the RAG Triad metric in the RAGOps framework has improved the performance of the railway KMS chatbot.

Keywords— Knowledge Management System, RAGOps, LLM-as-a-Judge, RAG Triad, Railways