

BAB I

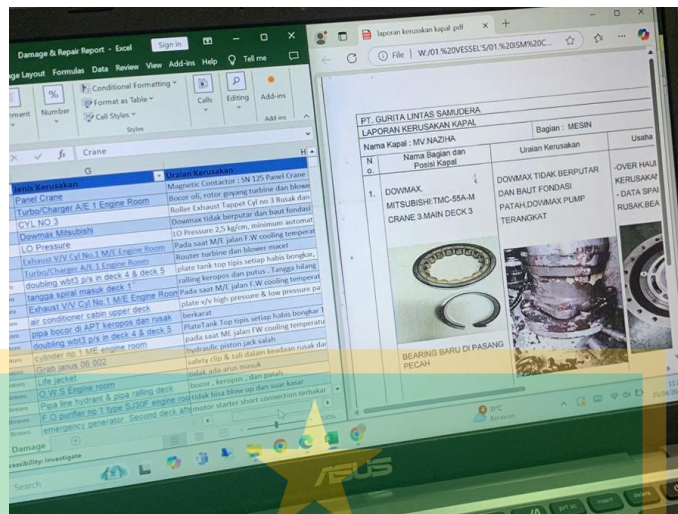
PENDAHULUAN

1.1 Latar Belakang

PT Pelayaran Gurita Lintas Samudera merupakan perusahaan yang bergerak di bidang pelayaran dan operasional kapal, yang secara rutin menghasilkan laporan kerusakan dan perbaikan kapal sebagai bagian dari kegiatan teknis dan administrasi. Laporan tersebut memuat informasi penting seperti nama kapal, tanggal kejadian, sistem atau komponen yang mengalami kerusakan, deskripsi kronologis, serta tindakan perbaikan yang dilakukan. Seiring dengan meningkatnya intensitas operasional kapal, jumlah laporan kerusakan yang dihasilkan setiap tahun juga mengalami peningkatan, sehingga volume data yang harus dikelola menjadi semakin besar dan kompleks.

Berdasarkan hasil observasi yang dilakukan pada bagian teknis dan administrasi kapal, proses pencarian laporan kerusakan masih dilakukan secara manual melalui file spreadsheet dan dokumen arsip yang tersimpan dalam berbagai folder terpisah. Ketika diperlukan informasi spesifik, seperti riwayat kerusakan kapal tertentu pada bulan dan tahun tertentu, staf harus membuka dan memeriksa dokumen satu per satu. Proses ini tidak hanya memakan waktu, tetapi juga berpotensi menimbulkan kesalahan identifikasi data serta keterlambatan dalam penyusunan laporan manajerial. Selain itu, sistem yang ada belum memiliki mekanisme pemeringkatan dokumen berdasarkan tingkat relevansi, sehingga hasil pencarian tidak terurut secara sistematis.

Kondisi penyimpanan dan pencarian laporan kerusakan kapal tersebut dapat dilihat pada Gambar 1.1.



Gambar 1. 1 Ilustrasi Penyimpanan Laporan Kerusakan Kapal dalam Format Berbeda

Pada sisi kiri terlihat data laporan dalam bentuk spreadsheet Microsoft Excel yang memuat informasi ringkas seperti jenis kerusakan dan lokasi mesin. Sementara itu, pada sisi kanan terlihat laporan rinci dalam bentuk dokumen PDF yang berisi uraian teknis serta dokumentasi visual hasil inspeksi. Perbedaan format dan lokasi penyimpanan ini menunjukkan bahwa data tidak terintegrasi dalam satu sistem terpadu. Akibatnya, proses pencarian harus dilakukan secara manual dengan membuka beberapa file berbeda, sehingga meningkatkan beban kerja serta menurunkan efisiensi pengelolaan informasi.

Dalam konteks Information Retrieval (IR), kualitas hasil pencarian sangat bergantung pada algoritma pemeringkatan dokumen yang digunakan. Sistem pencarian berbasis pencocokan kata kunci sederhana cenderung menghasilkan hasil yang tidak mempertimbangkan konteks dan tidak mampu mengurutkan dokumen berdasarkan tingkat relevansi terhadap query pengguna. Perkembangan penelitian terbaru menunjukkan bahwa pendekatan Retrieval-Augmented Generation (RAG) banyak digunakan untuk meningkatkan kualitas sistem pencarian berbasis dokumen dengan menggabungkan proses retrieval dan pemrosesan bahasa alami. Namun demikian, efektivitas sistem RAG sangat dipengaruhi oleh kualitas komponen retrieval yang digunakan.

Salah satu algoritma yang secara luas digunakan dan tetap menjadi baseline kuat dalam sistem retrieval modern adalah BM25 (Best Matching 25). BM25 merupakan model probabilistik yang menghitung skor relevansi dokumen dengan mempertimbangkan frekuensi kemunculan istilah dalam dokumen, inverse document frequency (IDF), serta normalisasi panjang dokumen. Pendekatan ini dinilai lebih stabil dibandingkan pembobotan klasik TF-IDF karena mampu mengurangi bias terhadap dokumen yang terlalu panjang dan lebih adaptif terhadap variasi struktur teks teknis (Cheng et al., 2025). Beberapa studi terbaru dalam arsitektur RAG juga menegaskan bahwa BM25 masih menjadi komponen retrieval yang efektif dalam sistem enterprise search dan knowledge-based retrieval (Gao et al., 2024).

Untuk meningkatkan akurasi pencarian berbasis konteks, sistem dalam penelitian ini juga mengintegrasikan Named Entity Recognition (NER) berbasis rule-based. NER berfungsi untuk mengekstraksi entitas penting dari query pengguna, seperti nama kapal, bulan, dan tahun, sebelum proses pemeringkatan dokumen dilakukan. Penerapan NER dalam domain teknis terbukti membantu meningkatkan ketepatan ekstraksi informasi pada dokumen pemeliharaan dan laporan kerja industri (Wei et al., 2025). Dengan adanya ekstraksi entitas tersebut, ruang pencarian dapat dipersempit sehingga hasil retrieval menjadi lebih terarah dan sesuai dengan kebutuhan pengguna.

Arsitektur sistem dikembangkan dalam kerangka Retrieval-Augmented Generation (RAG), di mana BM25 digunakan sebagai komponen utama dalam proses retrieval untuk mengambil dokumen yang paling relevan terhadap query pengguna. Dokumen yang telah diperingkat kemudian ditampilkan melalui antarmuka berbasis Streamlit yang dilengkapi dengan dashboard visualisasi untuk menyajikan tren kerusakan kapal secara informatif. Dalam penelitian ini, metode TF-IDF tetap digunakan pada tahap representasi teks dan pengolahan statistik guna mendukung kebutuhan agregasi data serta visualisasi pada dashboard, namun tidak digunakan sebagai metode evaluasi utama dalam pengujian performa retrieval.

Agar sistem yang dikembangkan tidak hanya bersifat implementatif tetapi juga terukur secara ilmiah, penelitian ini melakukan evaluasi performa modul retrieval menggunakan metrik Information Retrieval, yaitu Precision, Recall, dan F1-Score. Penggunaan metrik ini bertujuan untuk mengukur sejauh mana dokumen yang diperingkat oleh algoritma BM25 benar-benar relevan terhadap query pengguna, serta seberapa banyak dokumen relevan dalam dataset yang berhasil ditemukan sistem.

Precision digunakan untuk mengukur ketepatan sistem dalam menampilkan dokumen yang relevan, sedangkan Recall digunakan untuk menilai kemampuan sistem dalam menemukan seluruh dokumen yang relevan dalam koleksi data. F1-Score digunakan sebagai indikator keseimbangan antara keduanya. Evaluasi ini menjadi penting karena skor relevansi yang dihasilkan oleh algoritma BM25 saja belum cukup untuk menggambarkan kualitas sistem secara menyeluruh tanpa pengujian kuantitatif.

Dengan adanya evaluasi berbasis metrik tersebut, penelitian ini tidak hanya menunjukkan bahwa sistem mampu menghasilkan dokumen dengan skor relevansi tinggi, tetapi juga membuktikan secara terukur bahwa hasil pemeringkatan tersebut memiliki tingkat ketepatan dan kualitas yang baik. Pendekatan ini memperkuat kontribusi penelitian dalam ranah Information Retrieval pada domain industri pelayaran, yang masih relatif terbatas dibandingkan sektor lain seperti hukum, pendidikan, atau layanan pelanggan.

Selain itu, integrasi NER dan BM25 dalam arsitektur RAG memberikan pendekatan hybrid antara ekstraksi entitas berbasis aturan dan pemeringkatan dokumen berbasis probabilistik. Kombinasi ini memungkinkan sistem untuk memahami konteks permintaan pengguna secara lebih spesifik sebelum melakukan proses retrieval, sehingga hasil pencarian menjadi lebih terarah dan efisien. Hal ini sangat relevan dalam lingkungan operasional pelayaran, di mana kecepatan dan ketepatan akses informasi teknis dapat berdampak langsung pada keselamatan dan efektivitas operasional kapal.

Dengan latar belakang tersebut, penelitian ini diharapkan mampu memberikan solusi sistem pencarian yang tidak hanya meningkatkan efisiensi kerja, tetapi juga menyediakan kerangka evaluasi yang terukur dan dapat direplikasi untuk pengembangan sistem Information Retrieval pada domain industri lainnya.

1.2 Identifikasi Masalah

Berdasarkan hasil observasi dan analisis terhadap pengelolaan laporan kerusakan kapal di PT Pelayaran Gurita Lintas Samudera, ditemukan beberapa permasalahan utama yang menjadi dasar penelitian ini, yaitu:

1. Data laporan kerusakan kapal masih tersimpan dalam berbagai format yang tersebar dan belum terintegrasi dalam suatu sistem atau platform pencarian berbasis web.
2. Proses pencarian informasi dilakukan secara manual dengan membuka dan memeriksa dokumen satu per satu, sehingga kurang efisien dan berpotensi menimbulkan kesalahan identifikasi data.
3. Belum tersedia sistem pencarian yang mampu mengenali entitas penting seperti nama kapal, bulan, dan tahun secara otomatis berdasarkan query pengguna.
4. Belum terdapat mekanisme pemeringkatan dokumen yang dapat mengurutkan hasil pencarian berdasarkan tingkat relevansi terhadap kebutuhan pengguna dan evaluasi performa retrieval menggunakan metrik kuantitatif seperti Precision, Recall, dan F1-Score.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana merancang dan mengembangkan sistem intelligent search berbasis arsitektur Retrieval-Augmented Generation (RAG) untuk pencarian data kerusakan kapal di PT Pelayaran Gurita Lintas Samudera?
2. Bagaimana menerapkan metode Named Entity Recognition (NER)

berbasis rule-based untuk mengenali entitas penting seperti nama kapal, bulan, dan tahun dari query pengguna?

3. Bagaimana mengimplementasikan algoritma BM25 sebagai metode pemeringkatan dokumen untuk menghasilkan hasil pencarian yang relevan dan terurut dan mengevaluasi performa sistem retrieval menggunakan metrik Precision, Recall, dan F1-Score??
4. Bagaimana mengintegrasikan modul NER dan BM25 ke dalam platform web yang dilengkapi dengan dashboard visualisasi guna mendukung pencarian dan analisis data kerusakan kapal?

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang telah diidentifikasi, tujuan penelitian ini adalah:

1. Mengembangkan sistem Intelligent Search berbasis arsitektur Retrieval-Augmented Generation (RAG) untuk pencarian data kerusakan kapal pada PT Pelayaran Gurita Lintas Samudera.
2. Membangun model Named Entity Recognition (NER) berbasis rule-based yang mampu mengenali dan mengekstraksi entitas penting seperti nama kapal, bulan, dan tahun dari query pengguna.
3. Menerapkan algoritma BM25 sebagai metode pemeringkatan dokumen untuk mengurutkan hasil pencarian berdasarkan tingkat relevansi terhadap query pengguna dan Mengevaluasi performa sistem menggunakan metrik Precision, Recall, dan F1-Score..
4. Mengintegrasikan modul NER dan BM25 ke dalam platform web berbasis Streamlit yang dilengkapi dengan dashboard visualisasi untuk mendukung analisis data kerusakan kapal.

1.5 Batasan Masalah

1. Data yang digunakan terbatas pada laporan kerusakan kapal PT Pelayaran Gurita Lintas Samudera yang telah dikonversi kedalam format terstruktur untuk kebutuhan pengelolaan sistem.

2. Model Named Entity Recognition (NER) yang digunakan merupakan pendekatan rule-based berbasis vocabulary matching dan regex sederhana untuk mengenali entitas seperti nama kapal, bulan dan tahun dari query pengguna.
3. Pengujian performa modul retrieval pada arsitektur Retrieval-Augmented Generation (RAG) dilakukan menggunakan algoritma BM25 sebagai metode pemeringkatan dan penilaian relevansi dokumen terhadap query pengguna dan Evaluasi performa retrieval menggunakan metrik Precision, Recall, dan F1-Score..
4. Sistem yang dikembangkan difokuskan pada fungsi pencarian dan visualisasi data kerusakan kapal, serta tidak mencakup fitur prediksi kerusakan atau analisis berbasis machine learning lanjutan.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Manfaat Teknis
Menghasilkan sistem Intelligent Search berbasis integrasi Named Entity Recognition (NER) dan algoritma BM25 dalam arsitektur Retrieval-Augmented Generation (RAG) yang mampu meningkatkan relevansi dan kecepatan pencarian data kerusakan kapal.
2. Manfaat Praktis
Membantu PT Pelayaran Gurita Lintas Samudera dalam mempercepat proses pencarian laporan kerusakan kapal sehingga lebih efisien dibandingkan metode manual yang sebelumnya digunakan.
3. Manfaat Manajerial
Menyediakan dashboard visualisasi yang dapat membantu manajemen dalam memantau tren kerusakan kapal dan mendukung pengambilan keputusan berbasis data.
4. Manfaat Akademik

Menjadi referensi dalam pengembangan sistem Information Retrieval berbasis RAG dengan evaluasi kuantitatif menggunakan Precision, Recall, dan F1-Score pada domain industri pelayaran.

