

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kemajuan teknologi informasi yang semakin pesat telah mendorong perubahan besar pada cara perusahaan mengelola dan memanfaatkan data [1]. Volume data yang terus bertambah, baik dalam bentuk terstruktur maupun tidak terstruktur, menuntut adanya pendekatan analitik yang lebih canggih [2]. Pemanfaatan *data science* menjadi salah satu solusi utama karena mampu mengubah data mentah yang kompleks menjadi informasi bernilai untuk mendukung proses pengambilan keputusan strategis.

Sektor pertambangan batubara merupakan salah satu industri yang sangat bergantung pada ketersediaan data dalam berbagai bentuk [3]. Informasi tersebut mencakup dokumen teks yang berisi laporan internal maupun arsip eksternal, serta formulir inspeksi lapangan yang umumnya disimpan dalam bentuk hasil *scan*. Seiring dengan aktivitas operasional yang berjalan secara kontinu, jumlah dokumen yang dihasilkan dan disimpan semakin besar dari waktu ke waktu. Dalam pelaksanaan magang, korpus dokumen yang digunakan pada tahap *proof of concept* telah mencakup puluhan dokumen PDF sebagai bahan uji untuk kebutuhan pencarian informasi dan evaluasi awal. Kondisi ini menghadirkan tantangan tersendiri, yaitu bagaimana mengolah dan memanfaatkan data tersebut secara efektif agar dapat meningkatkan efisiensi kerja, menekan biaya operasional, serta menjaga daya saing di tengah kondisi pasar yang kompetitif [4]. Oleh karena itu, integrasi teknologi berbasis *artificial intelligence* (AI) dan *machine learning* (ML) menjadi semakin penting untuk mendukung transformasi digital di industri ini.

Salah satu tantangan yang dihadapi perusahaan berkaitan dengan pengelolaan dokumen teks dalam jumlah besar, baik berupa laporan internal maupun arsip eksternal seperti berita dan publikasi industri [5]. Proses pencarian informasi pada dokumen yang panjang sering kali membutuhkan waktu lama dan berpotensi menimbulkan kesalahan interpretasi apabila dilakukan secara manual [6]. Untuk menjawab tantangan tersebut, dikembangkan sistem berbasis *Large Language Model* (LLM) dengan pendekatan *Retrieval-Augmented Generation* (RAG). Sistem ini memungkinkan pengguna untuk mengunggah dokumen, kemudian memperoleh jawaban atas pertanyaan yang diajukan berdasarkan isi

dokumen. Pendekatan ini diharapkan dapat mempercepat proses pencarian informasi serta meningkatkan akurasi dalam mengekstraksi wawasan yang relevan bagi kebutuhan perusahaan.

Selain pengelolaan dokumen teks, pengolahan data operasional lapangan juga menghadapi kendala tersendiri [7]. Proses inspeksi peralatan dan aktivitas harian tambang umumnya dilakukan menggunakan formulir cetak yang berisi indikator kondisi peralatan dalam bentuk *bubble*, yaitu lingkaran yang diisi untuk menunjukkan kondisi baik atau tidak baik. Formulir tersebut kemudian disimpan sebagai dokumen hasil *scan* [8]. Proses *input* data dari formulir ini ke dalam sistem perusahaan masih banyak dilakukan secara manual. Kondisi tersebut berpotensi memakan waktu, menimbulkan ketidakkonsistenan pencatatan, serta membuka peluang terjadinya kesalahan [9], terutama ketika volume formulir yang harus diproses semakin meningkat.

Untuk mengatasi permasalahan tersebut, diterapkan teknologi *Optical Mark Recognition* (OMR) untuk mendeteksi *bubble* pada formulir hasil *scan*. Dengan teknologi ini, proses ekstraksi data dari formulir diharapkan dapat dilakukan secara otomatis, lebih cepat, dan konsisten, sehingga mampu meningkatkan efisiensi administrasi serta keandalan data operasional. Namun, sistem yang dikembangkan masih berada pada tahap *proof of concept* dan belum terintegrasi secara langsung dengan sistem perusahaan. Oleh karena itu, diperlukan pengujian model dan evaluasi performa untuk memastikan bahwa solusi yang dibangun benar-benar sesuai dengan kebutuhan operasional perusahaan sebelum dilakukan integrasi lebih lanjut.

Dengan demikian, pengembangan sistem RAG untuk menjawab pertanyaan berbasis dokumen dan implementasi OMR untuk deteksi *bubble* pada formulir hasil *scan* menjadi bagian penting dari penerapan teknologi *data science* dalam mendukung kebutuhan perusahaan. Kedua pendekatan tersebut diarahkan untuk menjawab tantangan pengelolaan data dalam jumlah besar, meningkatkan efisiensi proses kerja, serta menyediakan dasar yang andal bagi pengambilan keputusan di lingkungan industri pertambangan batubara.

1.2 Maksud dan Tujuan Kerja Magang

Maksud dari pelaksanaan kerja magang adalah sebagai berikut:

1. Memberikan pengalaman praktis dalam penerapan konsep *data science* pada konteks industri pertambangan.

2. Menambah wawasan tentang penerapan teknologi digital untuk mendukung efisiensi proses bisnis perusahaan.
3. Mengembangkan kompetensi profesional melalui keterlibatan langsung dalam penyelesaian permasalahan berbasis data di perusahaan.

Tujuan dari pelaksanaan kerja magang ini adalah sebagai berikut:

1. Mengembangkan sistem LLM berbasis RAG untuk ekstraksi informasi dari dokumen.
2. Mengembangkan sistem OMR untuk mendeteksi *bubble* pada formulir hasil *scan*.

1.3 Waktu dan Prosedur Pelaksanaan Kerja Magang

Kerja magang di PT Berau Coal Energy dilaksanakan selama enam bulan, mulai dari 17 Juni 2025 hingga 15 Desember 2025, dengan sistem kerja *hybrid* yang mengombinasikan *Work From Home* (WFH) dan *Work From Office* (WFO). Jadwal kegiatan mengikuti ketentuan perusahaan, yaitu Senin hingga Kamis pukul 08.30–17.45 WIB dan Jumat pukul 08.30–17.15 WIB. WFH diberlakukan pada hari Senin, Selasa, Rabu, dan Jumat, sedangkan hari Kamis ditetapkan sebagai WFO di kantor PT Berau Coal Energy, Sinarmas MSIG Tower lantai 10, Jl. Jend. Sudirman Kav. 21, Karet, Setiabudi, Jakarta Selatan. Selama WFH, kegiatan dilakukan secara daring melalui platform Zoom, sementara presensi dilakukan melalui dokumen Microsoft Word yang diisi mandiri dan diverifikasi oleh *supervisor* perusahaan.