

**PENGEMBANGAN SISTEM LLM BERBASIS
RETRIEVAL-AUGMENTED GENERATION DAN
SISTEM OPTICAL MARK RECOGNITION DI PT
BERAU COAL ENERGY**



LAPORAN CAREER ACCELERATION PROGRAM

**REVAN REVIAN
00000067662**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**PENGEMBANGAN SISTEM LLM BERBASIS
RETRIEVAL-AUGMENTED GENERATION DAN
SISTEM OPTICAL MARK RECOGNITION DI PT
BERAU COAL ENERGY**



LAPORAN CAREER ACCELERATION PROGRAM

**REVAN REVIAN
00000067662**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Revan Revian

NIM : 00000067662

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan MBKM Magang saya yang berjudul:

Pengembangan Sistem LLM Berbasis Retrieval-Augmented Generation dan Sistem Optical Mark Recognition di PT Berau Coal Energy

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 29 Desember 2025



(Revan Revian)

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Revan Revian
NIM : 00000067662
Program Studi : Informatika
Judul Laporan : Pengembangan Sistem LLM Berbasis Retrieval-Augmented Generation dan Sistem Optical Mark Recognition di PT Berau Coal Energy

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan sebagai berikut (beri tanda centang yang sesuai):

- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☐ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 29 Desember 2025



(Revan Revian)

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama Lengkap : Revan Revian
NIM : 00000067662
Program Studi : Informatika
Fakultas : Teknik dan Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : PT Berau Coal Energy
Alamat : Sinarmas MSIG Tower, Lantai 10.
Jl. Jend. Sudirman Kav. 21.
RT 012 RW 001
Kel. Karet, Kec. Setiabudi
Jakarta Selatan 12920

Email Perusahaan/Organisasi : corporate.communication@beraucoal.co.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat divalidasi keberadaannya.
2. Jika di kemudian hari terbukti ditemukan kecurangan atau penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.
 - b. Saya bersedia menerima semua sanksi yang berlaku sebagaimana ditetapkan dalam peraturan yang berlaku di Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 29 Desember 2025

Mengetahui,

PT. Berau Coal Energy Tbk 

(Nadika Sinatrya)

Menyatakan,





(Revan Revian)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Revan Revian
NIM : 00000067662
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Pengembangan Sistem LLM Berbasis Retrieval-Augmented Generation dan Sistem Optical Mark Recognition di PT Berau Coal Energy

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk memublikasikan hasil laporan saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh civitas academica UMN/publik. Saya menyatakan bahwa laporan yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia memublikasikan hasil laporan ini ke dalam repositori Knowledge Center karena ada data yang bersifat konfidensial.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara.
 - ☐ Embargo publikasi laporan dalam kurun waktu 3 tahun.

Tangerang, 29 Desember 2025

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Revan Revian

Halaman Persembahan / Motto

“Unless the Lord builds the house, They labor in vain who build it.”

Psalms 127:1 (NASB)



KATA PENGANTAR

Puji syukur dipanjatkan ke hadirat Tuhan Yang Maha Esa atas terselesainya kegiatan magang di PT Berau Coal Energy serta penyusunan laporan berjudul “Pengembangan Sistem *Retrieval-Augmented Generation* dan Sistem *Optical Mark Recognition* di PT Berau Coal Energy” sebagai salah satu syarat kelulusan Program Strata 1 Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara. Penyusunan laporan ini dapat terlaksana melalui dukungan dari berbagai pihak, sehingga ucapan terima kasih disampaikan kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Wirawan Istiono, S.Kom., M.Kom., selaku pembimbing laporan yang telah memberikan arahan dan masukan.
5. Bapak Nadika Sinatrya, selaku pembimbing magang di PT Berau Coal Energy.
6. Pihak-pihak lain yang telah berkontribusi dalam bentuk dukungan moral maupun material.

Harapannya, laporan ini dapat memberikan manfaat serta menjadi rujukan bagi pengembangan sistem serupa di masa mendatang.

Tangerang, 29 Desember 2025



Revan Revian

**PENGEMBANGAN SISTEM LLM BERBASIS
RETRIEVAL-AUGMENTED GENERATION DAN SISTEM OPTICAL
MARK RECOGNITION DI PT BERAU COAL ENERGY**

Revan Revian

ABSTRAK

Perkembangan pemanfaatan *data science* di industri pertambangan batubara mendorong kebutuhan sistem yang mampu mengolah dokumen teks dan formulir hasil *scan* secara otomatis. Di PT Berau Coal Energy, dua permasalahan utama diidentifikasi, yaitu lamanya pencarian informasi pada dokumen panjang serta tingginya beban *input* manual dari formulir inspeksi lapangan. Untuk mengatasi hal tersebut, dikembangkan dua sistem. Pertama, sistem *Large Language Model* (LLM) berbasis *Retrieval-Augmented Generation* (RAG) yang memanfaatkan *Qwen/Qwen3-Embedding-0.6B* untuk *embedding*, *jinaai/jina-reranker-v2-base-multilingual* untuk *reranking*, dan *Meta Llama 3 8B Instruct* untuk menghasilkan jawaban. Kedua, sistem *Optical Mark Recognition* (OMR) berbasis YOLO dan *TinyPatchCNN* untuk mendeteksi dan mengklasifikasi *bubble* pada formulir inspeksi. Model OMR dilatih menggunakan *synthetic dataset* dan disesuaikan melalui *fine-tuning* dengan *real scanned crops*. Hasil pengembangan divalidasi melalui pemeriksaan kualitatif pada sistem RAG dan inspeksi visual pada *output* OMR, sehingga mendukung ekstraksi informasi yang lebih cepat, konsisten, dan siap diintegrasikan ke *workflow* operasional.

Kata kunci: *Data science, large language model, optical mark recognition, retrieval-augmented generation*

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

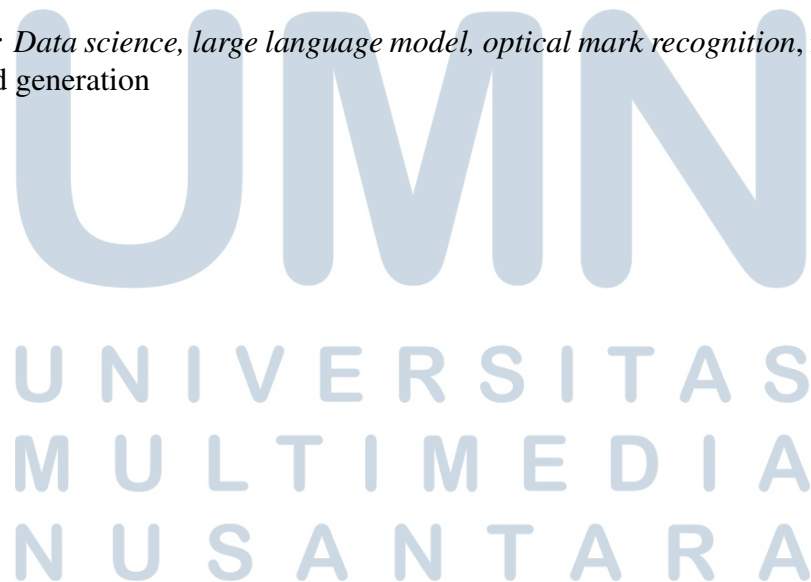
**DEVELOPMENT OF LLM-BASED RETRIEVAL-AUGMENTED
GENERATION AND OPTICAL MARK RECOGNITION SYSTEMS AT PT
BERAU COAL ENERGY**

Revana Reviana

ABSTRACT

The growing adoption of data science in the coal mining industry drives the need for systems capable of automatically processing text documents and scanned inspection forms. At PT Berau Coal Energy, two main challenges were identified, namely the time-consuming retrieval of information from lengthy documents and the high manual input workload associated with field inspection forms. To address these challenges, two systems were developed. The first is a Large Language Model (LLM) system based on Retrieval-Augmented Generation (RAG), utilizing Qwen/Qwen3-Embedding-0.6B for embedding, jinaai/jina-reranker-v2-base-multilingual for reranking, and Meta Llama 3 8B Instruct for answer generation. The second is an Optical Mark Recognition (OMR) system built with YOLO and TinyPatchCNN to detect and classify bubbles on inspection forms. The OMR models were trained using synthetic datasets and further refined through fine-tuning with real scanned crops. The developed systems were validated through qualitative review for the RAG system and visual inspection of OMR outputs, supporting faster and more consistent information extraction that is ready to be integrated into operational workflows.

Keywords: *Data science, large language model, optical mark recognition, retrieval-augmented generation*

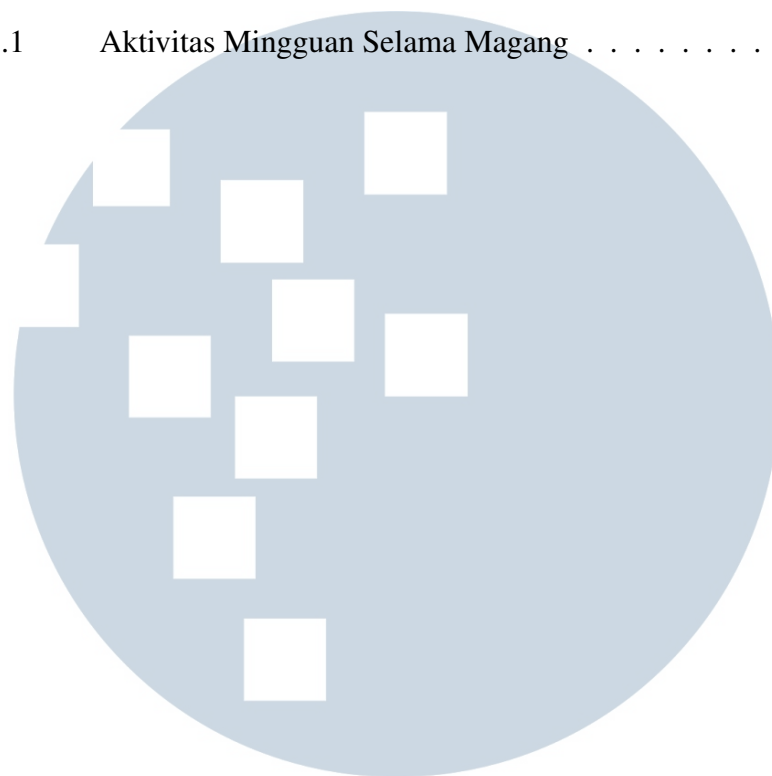


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Maksud dan Tujuan Kerja Magang	2
1.3 Waktu dan Prosedur Pelaksanaan Kerja Magang	3
BAB 2 GAMBARAN UMUM PERUSAHAAN	4
2.1 Deskripsi Singkat Perusahaan	4
2.2 Sejarah Singkat Perusahaan	5
2.3 Visi dan Misi Perusahaan	6
2.4 Struktur Organisasi Perusahaan	6
BAB 3 PELAKSANAAN KERJA MAGANG	8
3.1 Kedudukan dan Koordinasi	8
3.2 Tugas yang Dilakukan	9
3.3 Uraian Pelaksanaan Magang	9
3.4 Pengembangan Sistem LLM Berbasis RAG untuk Ekstraksi Informasi dari Dokumen	10
3.5 Pengembangan Sistem <i>Optical Mark Recognition</i> untuk Deteksi <i>Bubble</i> pada Formulir Hasil <i>Scan</i>	17
3.6 Kendala dan Solusi yang Ditemukan	30
BAB 4 SIMPULAN DAN SARAN	31
4.1 Simpulan	31
4.2 Saran	31
DAFTAR PUSTAKA	32

DAFTAR TABEL

Tabel 3.1	Aktivitas Mingguan Selama Magang	9
-----------	--	---



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1	Struktur organisasi Divisi CKBE PT Berau Coal Energy . .	6
Gambar 3.1	Struktur organisasi <i>Mining Geology & AI Department</i> pada divisi CKBE	8
Gambar 3.2	<i>Pipeline</i> sistem RAG untuk ekstraksi informasi dari dokumen internal	11
Gambar 3.3	Tampilan antarmuka untuk <i>upload</i> dokumen PDF pada tahap <i>ingestion</i>	12
Gambar 3.4	Status <i>vector store</i> setelah proses <i>embedding</i> dan pengembangan indeks <i>FAISS</i> berhasil dilakukan	14
Gambar 3.5	Antarmuka utama chatbot berbasis RAG pada tahap <i>generation post-processing</i>	16
Gambar 3.6	<i>Pipeline</i> arsitektur OMR yang digunakan untuk deteksi dan klasifikasi <i>bubble</i> pada formulir hasil <i>scan</i>	18
Gambar 3.7	Ilustrasi pembagian baris dan <i>region of interest</i> (ROI) untuk setiap pertanyaan pada formulir inspeksi.	20
Gambar 3.8	Contoh halaman sintetis yang digunakan untuk melatih pendeteksi YOLO	22
Gambar 3.9	Ringkasan arsitektur YOLO yang digunakan sebagai detektor pada tahap <i>synthetic training</i>	23
Gambar 3.10	<i>Log training</i> YOLO pada <i>dataset</i> sintetis. <i>Log</i> menampilkan perkembangan <i>loss</i> dan mAP pada setiap epoch	24
Gambar 3.11	Contoh <i>crop</i> sintetis untuk melatih <i>patch-level classifier</i> . .	24
Gambar 3.12	<i>Log training</i> TinyPatchCNN pada dataset sintetis. <i>Log</i> menunjukkan perkembangan <i>loss</i> dan akurasi di setiap <i>epoch</i>	25
Gambar 3.13	Contoh <i>real scanned crops</i> untuk proses <i>fine-tuning</i> TinyPatchCNN	25
Gambar 3.14	Contoh hasil <i>inference</i> yang menampilkan bounding box berwarna sesuai label <i>Baik</i> , <i>Tidak Baik</i> , dan <i>Kosong</i>	28
Gambar 3.15	Contoh tabel hasil <i>inference</i> yang diekspor ke format CSV	29

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	PRO-STEP-01 Cover Letter PRO-STEP Career Acceleration Program Track 2	33
Lampiran 2	PRO-STEP 02 PRO-STEP Career Acceleration Program Track 2 Card	34
Lampiran 3	PRO-STEP 03 Daily Task - PRO-STEP Career Acceleration Program Track 2	35
Lampiran 4	PRO-STEP 04 Verification Form of Internship Report PROSTEP Career Acceleration Program Track 2	46
Lampiran 5	Form Bimbingan	47
Lampiran 6	Hasil Pengecekan Similarity Turnitin	48
Lampiran 7	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	52

