

**PENERAPAN DATA SCIENCE UNTUK ANALISIS TEKS,
PREDIKSI HARGA KOMODITAS, DAN
IDENTIFIKASI ZONA BATUBARA
DI PT BERAU COAL ENERGY**



LAPORAN MBKM MAGANG

**REVAN REVIAN
00000067662**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**PENERAPAN DATA SCIENCE UNTUK ANALISIS TEKS,
PREDIKSI HARGA KOMODITAS, DAN
IDENTIFIKASI ZONA BATUBARA
DI PT BERAU COAL ENERGY**



**REVAN REVIAN
00000067662**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

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:

Penerapan Data Science untuk Analisis Teks, Prediksi Harga Komoditas, dan Identifikasi Zona Batubara 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, 20 Juni 2025



(Revan Revian)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Revan Revian
NIM : 00000067662
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Laporan MBKM Magang

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- ☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 20 Juni 2025

Yang menyatakan,



Revan Revian

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan kegiatan magang di PT Berau Coal Energy serta menyusun laporan berjudul “Penerapan Data Science untuk Analisis Teks, Prediksi Harga Komoditas, dan Identifikasi Zona Batubara 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 tentu tidak terlepas dari bantuan dan dukungan berbagai pihak, baik secara moral maupun material. Atas segala kontribusi tersebut, ucapan terima kasih yang sebesar-besarnya 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 Dennis Gunawan, S.Kom., M.Sc., sebagai dosen pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan dan arahan.
5. Bapak Nadika Sinatrya sebagai pembimbing selama pelaksanaan magang di PT Berau Coal Energy yang telah memberikan arahan, dukungan, dan masukan selama menjalani kegiatan magang.
6. Seluruh pihak yang telah membantu dalam penulisan laporan magang ini, baik secara langsung maupun tidak langsung.

Tangerang, 20 Juni 2025



Revan Revian

PENERAPAN DATA SCIENCE UNTUK ANALISIS TEKS, PREDIKSI HARGA KOMODITAS, DAN IDENTIFIKASI ZONA BATUBARA DI PT BERAU COAL ENERGY

Revan Revian

ABSTRAK

Perkembangan teknologi digital mendorong penerapan *data science* dalam industri pertambangan batubara untuk meningkatkan efisiensi operasional dan kualitas pengambilan keputusan. Di PT Berau Coal Energy, tiga proyek utama telah dikembangkan. Pertama, sistem analisis teks berbasis LLM digunakan untuk mengolah artikel berita daring melalui proses *scraping*, *summarization*, *sentiment analysis*, dan analisis risiko-peluang. Proyek ini memanfaatkan model *bart-large-cnn* untuk *summarization*, *financial-roberta-large-sentiment* untuk *sentiment analysis*, dan *llama-3-8b-Instruct-bnb-4bit* untuk menghasilkan respons analisis risiko dan peluang secara kontekstual. Kedua, prediksi harga komoditas dilakukan dengan teknik *feature engineering* berbasis *time series*, membandingkan algoritma statistik, *machine learning*, dan *deep learning*. Model GRU memberikan hasil terbaik dengan MAE 145,93, RMSE 224,42, dan MAPE 1,56%. Ketiga, pendekatan *clustering* dengan GMM diterapkan pada data *well logging* geofisika untuk mengidentifikasi zona batubara; ditemukan lima *cluster* optimum, dengan *Cluster 0* menunjukkan nilai *Coal Index* tertinggi, tervalidasi melalui PCA dan t-SNE. Ketiga proyek ini menunjukkan peran penting analitik modern dalam mempercepat pemantauan tren, meningkatkan akurasi prediksi pasar, dan memfokuskan eksplorasi.

Kata kunci: Analisis teks, *data science*, identifikasi zona batubara, prediksi harga komoditas, PT Berau Coal Energy

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

**IMPLEMENTATION OF DATA SCIENCE FOR TEXT ANALYSIS,
COMMODITY PRICE FORECASTING, AND COAL ZONE
IDENTIFICATION AT PT BERAU COAL ENERGY**

Revan Revian

ABSTRACT

The advancement of digital technology has driven the adoption of data science in the coal mining industry to improve operational efficiency and decision-making quality. At PT Berau Coal Energy, three main projects have been developed. First, a text analysis system based on LLMs was used to process online news articles through scraping, summarization, sentiment analysis, and risk-opportunity assessment. The project utilized facebook/bart-large-cnn for summarization, financial-roberta-large-sentiment for sentiment analysis, and llama-3-8b-Instruct-bnb-4bit for generating contextual responses to risk and opportunity analysis. Second, commodity price forecasting was conducted using time series-based feature engineering, comparing statistical, machine learning, and deep learning algorithms. The GRU model delivered the best performance with an MAE of 145.93, RMSE of 224.42, and MAPE of 1.56%. Third, a clustering approach using GMM was applied to geophysical well logging data to identify coal zones; five optimal clusters were found, with Cluster 0 showing the highest Coal Index value, validated through PCA and t-SNE. These three projects demonstrate the strategic role of modern analytics in accelerating trend monitoring, enhancing market prediction accuracy, and focusing exploration efforts.

Keywords: Coal zone identification, commodity price prediction, data science, PT Berau Coal Energy, text analysis

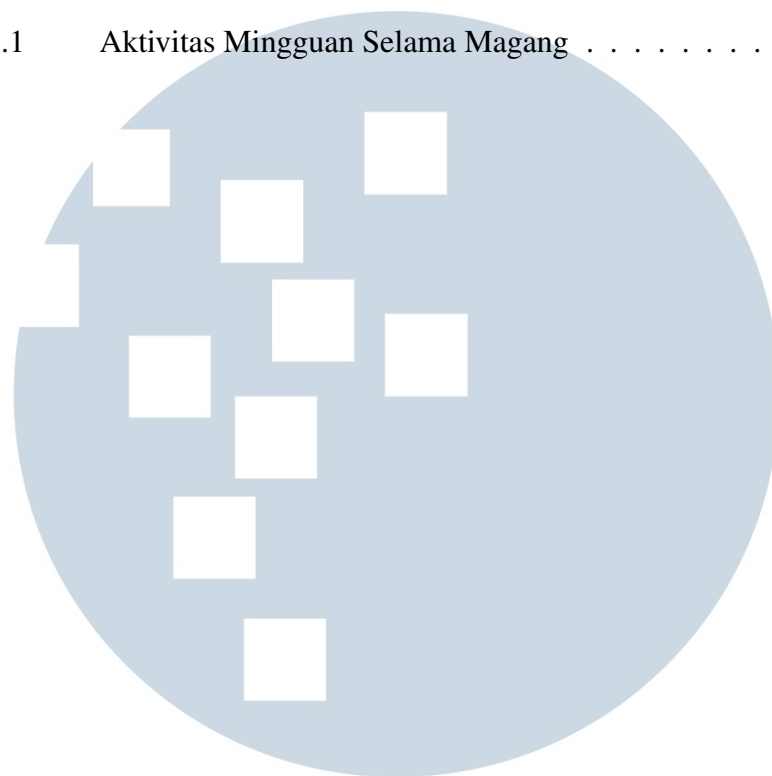
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
HALAMAN PERSEMBAHAN/MOTO	iv
KATA PENGANTAR	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xii
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	10
3.3.1 Implementasi LLM untuk Analisis Artikel Berita Daring Terkait Perusahaan Tambang Global	10
3.3.2 Prediksi Harga Komoditas Menggunakan Model <i>Machine Learning</i> Berbasis Data <i>Time Series</i>	20
3.3.3 Membangun Model <i>Machine Learning</i> untuk Mengelompokkan Data <i>Logging</i> Geofisika dalam Rangka Identifikasi Zona Batubara	32
3.4 Kendala yang Ditemukan	49
3.5 Solusi atas Kendala yang Ditemukan	50
BAB 4 SIMPULAN DAN SARAN	51
4.1 Simpulan	51
4.2 Saran	51
DAFTAR PUSTAKA	53

DAFTAR TABEL

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



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 2.1	Struktur organisasi perusahaan PT Berau Coal Energy Divisi CKBE	6
Gambar 3.1	Struktur organisasi Mining Geology & AI Department pada divisi CKBE	8
Gambar 3.2	<i>Pipeline</i> untuk menganalisis artikel berita daring terkait pertambangan menggunakan LLM	11
Gambar 3.3	Contoh hasil proses <i>web scraping</i> terhadap artikel berita daring	12
Gambar 3.4	Contoh teks artikel sebelum dan sesudah tahap <i>preprocessing</i> dilakukan	13
Gambar 3.5	Contoh perbandingan hasil <i>summarization</i> dari beberapa model pada artikel yang sama	15
Gambar 3.6	Ringkasan evaluasi kuantitatif menggunakan metrik ROUGE dan BERTScore	15
Gambar 3.7	Rata-rata <i>confidence score</i> dari seluruh model <i>sentiment analysis</i> yang diuji pada dataset artikel	17
Gambar 3.8	Contoh hasil inferensi model-model LLM dalam mengidentifikasi risiko dan peluang dari satu artikel yang sama	19
Gambar 3.9	Contoh hasil akhir dalam format CSV yang berisi <i>summarization</i> , <i>sentiment analysis</i> , dan analisis risiko-peluang dari model terbaik	20
Gambar 3.10	<i>Pipeline</i> untuk memprediksi harga komoditas dengan memanfaatkan fitur berbasis <i>time series</i>	21
Gambar 3.11	Pergerakan harga komoditas mingguan dari waktu ke waktu	22
Gambar 3.12	Distribusi harga mingguan komoditas	23
Gambar 3.13	Dekomposisi musiman terhadap harga mingguan komoditas	23
Gambar 3.14	Plot ACF terhadap harga mingguan komoditas	24
Gambar 3.15	Plot PACF terhadap harga mingguan komoditas	24
Gambar 3.16	Hasil <i>dataframe</i> setelah proses <i>feature engineering</i>	26
Gambar 3.17	Visualisasi <i>feature importance</i> berdasarkan hasil Random Forest	27
Gambar 3.18	Tabel hasil evaluasi metrik MAE, RMSE, dan MAPE dari seluruh model	29
Gambar 3.19	Kurva perbandingan harga aktual dan prediksi lima model terbaik pada periode <i>testing</i> (Juli–Desember 2024)	30
Gambar 3.20	Kurva perbandingan harga aktual dan prediksi lima model terbaik pada periode <i>forecasting</i> (Januari–April 2025)	32
Gambar 3.21	<i>Pipeline</i> untuk mengelompokkan zona batubara berdasarkan data <i>geophysical logging</i>	34
Gambar 3.22	Distribusi masing-masing fitur utama (GR, LD, CL, SD) pada data geofisika	35
Gambar 3.23	Jumlah <i>missing values</i> pada masing-masing fitur utama	35
Gambar 3.24	Matriks korelasi antar fitur utama pada data geofisika	36
Gambar 3.25	Distribusi jumlah data pada masing-masing kelas litologi	36

Gambar 3.26	Contoh tampilan data hasil pembersihan data dan <i>feature engineering</i>	37
Gambar 3.27	Daftar fitur hasil seleksi setelah proses VIF <i>filtering</i>	38
Gambar 3.28	Evaluasi Random Forest pada data <i>training (in-sample)</i>	39
Gambar 3.29	Evaluasi XGBoost pada data <i>training (in-sample)</i>	40
Gambar 3.30	Evaluasi Random Forest pada data <i>testing (hold-out)</i>	40
Gambar 3.31	Evaluasi XGBoost pada data <i>testing (hold-out)</i>	40
Gambar 3.32	Evaluasi Random Forest pada label batubara orisinal (<i>full-data</i>)	41
Gambar 3.33	Evaluasi XGBoost pada label batubara orisinal (<i>full-data</i>)	41
Gambar 3.34	Jumlah interval batubara aktual yang berhasil dideteksi masing-masing model	41
Gambar 3.35	<i>Feature importance</i> pada model Random Forest dan XGBoost (<i>Top 10</i>)	42
Gambar 3.36	Visualisasi <i>log</i> geofisika, probabilitas model, dan label litologi pada satu sumur	43
Gambar 3.37	Perbandingan nilai BIC dan AIC untuk berbagai jumlah <i>cluster</i> pada GMM	44
Gambar 3.38	Distribusi jumlah data pada masing-masing <i>cluster</i>	45
Gambar 3.39	Rata-rata <i>Coal Index</i> dan proporsi <i>high-coal</i> pada tiap <i>cluster</i>	45
Gambar 3.40	Distribusi <i>Coal Index</i> terhadap kedalaman pada masing-masing <i>cluster</i>	46
Gambar 3.41	Distribusi <i>cluster</i> pada ruang dua komponen utama PCA	47
Gambar 3.42	Distribusi <i>Coal Index</i> pada ruang dua komponen utama PCA	47
Gambar 3.43	<i>Top 10 PCA loadings</i> untuk dua komponen utama (PC1 dan PC2)	48
Gambar 3.44	Visualisasi <i>cluster</i> pada ruang t-SNE	48
Gambar 3.45	Visualisasi <i>Coal Index</i> pada ruang t-SNE	49



DAFTAR LAMPIRAN

Lampiran 1	MBKM-01 Cover Letter MBKM Internship Track 1	54
Lampiran 2	MBKM-02 MBKM Internship Track 1 Card	55
Lampiran 3	MBKM-03 Daily Task - Internship Track 1	56
Lampiran 4	MBKM-04 Verification Form of Internship Report MBKM Internship Track 1	74
Lampiran 5	Form Bimbingan	75
Lampiran 6	Hasil Pengecekan Turnitin	76

