Fulandi Hudza Grahitama

Bogor, Indonesia | fulandihudza@gmail.com | +62 877-8157-4213 | LinkedIn | GitHub

SUMMARY

Informatics Engineering graduate (GPA 3.57/4.00) targeting Data Engineer roles. Hands-on with batch pipelines orchestrated by Airflow and Kestra (Python, Docker, PostgreSQL), dbt models on BigQuery, and Terraform on GCP.

TECHNICAL SKILLS

Data Engineering	:	Batch ETL/ELT pipelines, workflow orchestration (Airflow, Kestra), dbt modeling, data quality checks, deduplicated loads (MERGE), partitioned tables.
Programming	:	Python (pandas, NumPy, SQLAlchemy, click), SQL (PostgreSQL, BigQuery).
Cloud & Tools	:	GCP (Cloud Storage, BigQuery), Terraform, Docker, Docker Compose, Git, uv
Analytics & ML	:	Exploratory data analysis, scikit-learn (K-Means, PCA), UMAP, deep learning for satellite imagery (U-Net, ResUNet, ConvMAE), Streamlit, pytest, Tableau, Looker Studio, SPSS.
Languages	:	Indonesian (native), English (intermediate, TOEIC Prediction Test 715, Jul 2026)

EXPERIENCE

Data Engineer Intern at CarbonEthics ***Oct 2025 – present***

- Automated a weekly pipeline (Excel, Google Apps Script, Gemini API) that summarizes carbon registry news from Verra (every Thursday) and Gold Standard (every Monday), used by the entire project development team.
- Ran experiments with U-Net, ResUNet, and a ConvMAE model (dense satellite embeddings) for land cover classification to identify the best approach for production.
- Studied foundation models and model distillation (AlphaEarth), and demonstrated RAG with a vector database to users for a knowledge management system.

PROJECTS

Weather Data Pipeline: OpenWeatherMap to BigQuery | GitHub ***Nov 2026***

Python, Airflow, dbt, BigQuery, Docker

- Built an Airflow DAG (every 6 hours) that loads weather data for 5 Indonesian cities into BigQuery, then runs dbt staging, dimension, and partitioned fact models with 24 passing tests.
- Added a data quality audit (freshness, completeness, range checks) scored 0-100 and logged to BigQuery; critical failures stop the pipeline. Traced 2 failed runs (of 16) to local network loss.

Batch Data Pipeline: NYC Taxi to PostgreSQL and BigQuery | GitHub ***Sep – Oct 2026***

Python, Docker, Kestra, Terraform, GCP

- Built a parameterized Python CLI (pandas, SQLAlchemy) that loads gzipped CSVs into PostgreSQL in 100,000-row chunks, containerized with Docker and run with Docker Compose alongside pgAdmin.
- Orchestrated Kestra flows that load taxi data into PostgreSQL and BigQuery (via Cloud Storage), using MD5 row hashes and MERGE to avoid duplicates, with monthly cron schedules; Terraform for the bucket and dataset.

ArchetypeXI: Football Player Archetype and Similarity Platform | GitHub ***Sep – Oct 2026***

Python, scikit-learn, UMAP, Parquet, Streamlit, pytest

- Processed per-90-minute stats for 1,827 players into 35 features (VIF filtering, scaling, Parquet) and clustered them into 5 playing-style archetypes with K-Means; built a cosine-similarity engine with pytest tests.
- Deployed a multi-page Streamlit app on Streamlit Community Cloud (https://archetype-xi.streamlit.app)

EDUCATION

UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL ***Sep 2022 – Sep 2026***

D4 (S.Tr.Kom.) Informatics Engineering | GPA: 3.57 / 4.00

- Relevant courses: Database, Data Mining, Cloud Computing, Artificial Intelligence, Probability and Statistics. Thesis: U-Net Multimodal with Grad-CAM Explainability for Land Use/Land Cover Classification.

PUBLICATIONS

Design and Implementation of a RESTful API-Based Point of Sale System

Nuansa Informatika, Vol. 19 No. 1 (2025), pp. 31–38, SINTA 5 | Supabase, Express.js, 3NF database | DOI

Deep Learning-Based Semantic Segmentation for Land Cover Mapping from Remote Sensing Imagery: A Systematic Literature Review

JTIIK, Universitas Brawijaya | SINTA 2 | Second author | Accepted Oct 2026 (in production)