

# DION RICKY SAPUTRA

East Java, Indonesia

📞 085155458181 📩 [dionrickysptr@gmail.com](mailto:dionrickysptr@gmail.com) 💬 [linkedin.com/in/dion-ricky](https://linkedin.com/in/dion-ricky) 🐾 [github.com/dion-ricky](https://github.com/dion-ricky)

## Education

### University of Brawijaya

*Bachelor in Computer and Information Science*

Aug. 2018 – Jul. 2022

Malang, East Java

## Experience

### Blibli

*Data Engineer*

Sep. 2022 – Present

Jakarta, Indonesia

- Take part in cost reduction initiative and achieved ~77% cost reduction for BigQuery storage. Implement SCD type 2 for recording historical data replacing legacy tables and achieve ~97% storage cost reduction from ~200 USD/mo to ~6.3 USD/mo
- Build and integrate data lineage tracker with pipeline executor (Airflow, Prefect, and dbt). Develop a feature to support column-level lineage for more granular metadata. Build a new Data Warehouse project to support the Blibli subsidiary using Airflow and dbt. Develop and maintain ETL pipeline to support BAU. Working with various data sources (Postgres, MongoDB, Parquet, Avro, Protobuf)

### Blibli

Jun. 2021 – Mar 2022

*Data Engineer Intern*

Jakarta, Indonesia

- Deployed and managed Airflow for running data pipeline. Designed and created data warehouse to store AppStore and PlayStore app reviews for the purpose of analysis and machine learning. Trained machine learning model for classifying review sentiment. Explored various ML model ranging from the traditional model to the neural network model using CNN and LSTM. The best score is achieved using mixture of both traditional model and NN model using simple voting ensemble mechanism. Analyzed, designed, and created a dashboard for an app review analysis

### Google DSC

Aug. 2020 – Aug. 2021

*Volunteer*

Malang, Indonesia

- Managing public community social media as graphic designer. Hosted a UI/UX Workshop as the main speaker attended by 100+ audiences

## Certifications

### Professional Data Engineer

GOOGLE CLOUD PLATFORM

Jan. 2023 – Jan. 2025

*View Certificate*

## Projects

### Streaming Data Pipeline | Java, Rust, Kafka

March 2025

- Built a data pipeline to process streaming data and calculate aggregations on them as they arrive.
- Wrote custom data producer in Rust to periodically publish message to Kafka to replicate real-world scenario.
- Analyzed financial data in real-time to calculate total transaction value for every hour window.

### Mini "Airflow" Clone | Python, Google Cloud, CI/CD

November 2022

- Developed an "Airflow" clone that runs in GCP Cloud Function.
- This project is focused on improving my Python skill by learning design patterns such as singleton and decorators.
- Deployed using Github workflow using custom steps and deployment configuration.

### Reviews Sentiment Analysis | Machine Learning, Python

January 2022

- Using machine learning to predict the sentiment of an app review.
- Trained and compared multiple machine learning models with different hyperparameter which resulted in several models with specific traits and characteristics.
- Combined all trained models using voting mechanism in order to offset the unique weakness of each model but still producing good prediction accuracy.

### Twitter Scrapper | Python, Selenium

January 2022

- Built a Python library to scrape tweets in user's Twitter profile.

## Technical Skills

**Languages:** Python, SQL, Java, Javascript, HTML & CSS, Rust

**Cloud:** Google Cloud Platform, AWS, Azure

**Technologies/Frameworks:** Docker, Linux, Kubernetes, Jenkins, Spark, Kafka