Hizkia Gunawan

hizkialeon89@gmail.com
linkedin.com/in/hizkiagunawan7
(+62) 81384645098
github.com/Hizkia7
Tangerang, Indonesia

PROFILE

Machine Learning Engineer and Data Scientist, delivering scalable ML solutions across fintech and manufacturing. Skilled in end-to-end model development, cloud deployment (GCP, AWS), and building data-driven systems like credit scoring, pricing engines, and fraud detection. Experienced with tools such as Vertex AI, Kubeflow, PyMC, DBT, and Airflow. Adept at designing event-driven pipelines and deploying containerized microservices with FastAPI and Streamlit.

EXPERIENCE

Machine Learning & Data Engineer

Jul 2024 - Present

Infinid Indonesia | Jakarta, Indonesia

1st Project: Built Appraisal Engine to Estimate Property Price

Enabled stakeholders to instantly estimate property values through a fully automated system, eliminating the need for human involvement and reducing the appraisal process time.

Data Acquisition & Engineering:

- Developed scalable web scrapers using **Scrapy Redis** and **BeautifulSoup** to collect property listings from multiple websites.
- Containerized the scraping jobs as Docker images and published them to **GCP Artifact Registry**.
- Deployed containers to **GCP Cloud Run** with scheduled executions to automate data collection.
- Stored scraped raw data in **BigQuery** for downstream processing.
- Designed event-driven data pipelines using **Cloud Log Router**, **Pub/Sub**, and **Cloud Functions** to trigger data transformations automatically when new data arrived.
- Transformed raw data into clean, structured **staging** and **fact tables** in BigQuery.

Model Development & Deployment:

- Trained and compared multiple regression models (**Linear Regression**, **Gradient Boosting**, **MLP Regressor**) using **Vertex AI** with **Kubeflow Pipelines**.
- Selected the best-performing model, registered it in **Vertex AI Model Registry**, and deployed it as a prediction endpoint.
- Built a **FastAPI** backend to serve predictions via the deployed ML model.
- Created an interactive frontend using **Streamlit** to simulate the appraisal system for end-users.
- Deployed both backend and frontend as Docker containers to **GCP Cloud Run**, ensuring scalability and ease of maintenance.

2nd Project: Developed 3C Credit Scoring to Assess Borrower Creditworthiness.

Streamlined the credit evaluation process by empowering analysts to automatically pre-filter high-potential borrowers based on comprehensive risk profiles, significantly reducing manual effort and enabling faster, more targeted loan officer outreach.

Data Engineering:

- Built two event-driven data pipelines to extract and transform **Character, Capacity** from credit history data and **Collateral** from asset data.
- Triggered using **Cloud Log Router**, **Pub/Sub**, and **Cloud Functions** for near real-time updates.

Model Development, Integration, Monitoring:

- Designed and deployed a **Bayesian Credit Scoring Model** using **PYMC**.
- Packaged model in a Docker container and deployed to **Cloud Run**, triggered automatically upon 3C data update.
- Delivered real-time borrower credit scores integrated with **Zoho CRM**.
- Persisted scoring results in **BigQuery** for backtesting and performance evaluation.
- Implemented periodic update mechanisms for dynamic adjustment of scoring weights based on historical performance.

3rd Project: Migrate the 3C Credit Scoring Pipeline from GCP to AWS Cloud Environment for Scalability and Cost Optimization.

Migrated to AWS to reduce compute costs and offering more granular resource control for model execution.

Data Engineering:

- Extracted and transformed JSON-based credit history into structured datasets.
- Stored and queried data with **AWS Athena**.
- Transformed data into 3C components using **DBT**, orchestrated with **AWS MWAA (Managed Workflows for Apache Airflow)**.

Model Deployment:

- Containerized PYMC credit scoring model and pushed to **AWS ECR**.
- Deployed and executed the model using **AWS Batch**, orchestrated via **Airflow Batch Operator** for on-demand scoring.

Senior Data Scientist

Jan 2023 - Jun 2024

Pepper Advantage | London, United Kingdom

1st Project: Macroeconomic Customer Segmentation (Dataiku with Snowflake)

Enabled dynamic customer segmentation based on macroeconomic regimes and identifies evolving customer behavior patterns across countries. This empowered the business to adapt strategies based on market conditions and customer trends.

Model Development & Deployment

- Built a **multiclass classification model** to identify macroeconomic regimes across multiple countries using macro-financial indicators.
- Developed a **customer segmentation model** to cluster users based on historical payment behaviors.
- Extracted large-scale structured data from **Snowflake** for training and inference.
- Deployed both models to **Dataiku Automation Node** for scheduled predictions and seamless integration into downstream processes.

Dashboard Development and Data Analysis:

- Designed and developed a robust **data pipeline** in **Snowflake** to seamlessly join **macroeconomic regime results** with **customer segmentation** data, ensuring data integrity and analytical consistency.
- Built interactive, insight-driven dashboards in **Tableau**, enabling stakeholders to visualize macroeconomic trends alongside segmented customer behaviors.
- Integrated Tableau dashboards directly with **Snowflake** data sources, providing real-time access to updated insights for strategic decision-making.

2nd Project: Dynamic Client Portfolio Pricing App (GCP Vertex AI)

Optimized loan pricing strategies by delivering a real-time recommendation system that personalizes offers based on customer risk and repayment capacity—enabling more competitive, data-driven pricing decisions.

Designed two core models:

- A **binary classification model** to predict the likelihood of a customer making a future payment.
- A **regression model** to estimate each customer's **loan repayment capacity**.Orchestrated data workflows using **BigQuery** for storage and **Kubeflow on Vertex AI** for end-to-end model training and prediction.

Implemented **event-driven automation** via **Pub/Sub** to trigger ML pipelines based on new data updates.

Created an intuitive, interactive user interface using **Streamlit**, enabling business users to simulate pricing scenarios in real time.

3rd Project: Desk Collection Prediction System (GCP Vertex AI)

Enabled desk collection teams to optimize call strategies by prioritizing debtors based on their likelihood to pay. Predictive models provided real-time insights on which customers should be contacted or re-contacted, significantly improving the efficiency and effectiveness of debt collection efforts.

- Designed and implemented two core models:
- **Pre-Call Model**: Predicts a customer's likelihood to make a payment using loan portfolio and historical repayment data.
- **After-Call Model**: Updates payment probability based on **telephony interaction data** captured from collection calls.Leveraged **BigQuery** as the centralized data warehouse for sourcing and preparing input features.

Built and orchestrated an end-to-end ML pipeline using **Kubeflow on GCP Vertex AI**, enabling continuous training and scalable batch predictions.

Enabled real-time automation through **Pub/Sub**, which triggers model execution whenever new loan or call data is ingested.

Data Scientist

Dec 2021 - Dec 2022

PT. Sharing Vision Indonesia | Jakarta, Indonesia

Fraud Detection & Data Platform Development (Bank Raya Indonesia & Allo Bank)

At Bank Raya Indonesia – Account Take Over (ATO) Fraud Detection Model

Strengthened fraud prevention capabilities and accelerated data-driven decision-making to detect and mitigate fraud as soon as possible.

- Developed a machine learning model to detect **Account Take Over (ATO)** fraud, where fraudsters gain unauthorized access to customer accounts.
- Leveraged customer **transaction behavior patterns** and **IMEI-based device fingerprinting** to flag suspicious activities.
- Deployed the model as a real-time scoring **Flask API**, enabling seamless integration with banking systems for live fraud prevention.**At Allo Bank – End-to-End Data Platform Architecture & ELT Pipeline**

Empowered business and operations teams with unified, reliable customer insights by building a complete ELT data platform.

- Designed and implemented an **end-to-end data pipeline** to centralize and transform customer data across systems.
- Built an **ELT workflow using Apache Airflow** to extract customer data from **MySQL**, load it into **PostgreSQL**, and transform it using **DBT (Data Build Tool)**.
- Enabled data-driven decision-making by visualizing cleaned and modeled data in **Apache Superset**, creating dashboards for operational and business analytics.

Research Project Assistant

Aug 2019 - Jan 2021

Taiwan Ministry Of Science & Technology | Taoyuan, Taiwan, Province of China

Manufacturing Optimization Using AI (Avary Holding & Kinsus Interconnect Technology)

At Avary Holding (Shenzhen) Co., Ltd. – Machine Parameter Optimization via PSO-ANN

- Developed an AI-based system to **optimize manufacturing machine parameters**, improving production efficiency and product quality.
- Implemented an **Artificial Neural Network (ANN) using Multilayer Perceptron** architecture, trained to model complex relationships between machine settings and output metrics.
- Utilized **Particle Swarm Optimization (PSO)** to fine-tune machine parameters for optimal performance.
- Built and trained the model using **TensorFlow Keras**, enabling scalable and reproducible experimentation.**At Kinsus Interconnect Technology Corp. – Machine Parameter Optimization via Firefly-ANN**
- Designed a similar optimization framework using **Firefly Algorithm** as the metaheuristic strategy to enhance the ANN model's output in a high-precision manufacturing environment.
- Applied **Multilayer Perceptron models** in TensorFlow Keras to simulate manufacturing dynamics, enabling intelligent decision-making in production line setup.
- Resulted in reduced material waste and improved throughput by identifying optimal machine configurations through AI-driven simulations.

Project Manager

Jul 2017 - Jan 2019

PT. Pangkal Multikarya | Jakarta, Indonesia

Project Leadership & Site Execution Management

- **Led end-to-end project execution**, ensuring timely delivery by developing and managing detailed **project timeline schedules** aligned with technical and operational requirements.
- Coordinated with the engineering team to **analyze shop drawings** and accurately **calculate material requirements**, ensuring cost-efficiency and material availability.
- Played a key role in **client negotiations**, preparing cost estimates and securing agreement on price offerings prior to project kickoff.
- Tracked and reported **monthly progress updates using S-Curve analysis**, providing clear visibility into project health and milestones.
- Conducted **on-site inspections** during installation phases to ensure compliance with design specifications and quality standards.

EDUCATION

Master of Science

2021

Yuan Ze University | Taoyuan, Taiwan, Province of China

Specialized in applying **Artificial Intelligence (AI)**, **Machine Learning**, and **Big Data Analytics** to improve manufacturing systems and industrial operations.

Gained hands-on experience in developing data-driven solutions to optimize **production efficiency**, **process quality**, and **equipment performance**.

Studied advanced topics including:

- **Smart Manufacturing Systems**
- **AI-powered Optimization Algorithms**
- **Predictive Maintenance using Machine Learning**Completed practical projects involving **neural network modeling**, **metaheuristic optimization (e.g., PSO, Firefly Algorithm)**, and **real-time data analysis** using tools such as **TensorFlow Keras, Python.**

Developed a strong foundation in integrating **technology, data, and decision-making** to drive digital transformation in manufacturing environments.

Bachelor of Engineering

2016

Tarumanagara University | Jakarta, Indonesia

Built a strong foundation in **industrial engineering principles** and **manufacturing management systems**, focusing on the integration of people, processes, and technology.

Studied key areas such as:

- **Production Planning & Control**
- **Operations Research & Optimization**
- **Quality Management (TQM, Six Sigma)**
- **Supply Chain & Logistics Management**
- **Facility Layout & Work System Design**Gained practical experience in analyzing and improving **production systems**, **process flows**, and **resource allocation** using industrial engineering tools and methodologies.

Developed strong analytical and problem-solving skills to support **efficiency improvement**, **cost reduction**, and **workflow optimization** in industrial settings.

SKILLS

Python | SQL | Data Modeling / Statistical Modeling | Data Warehousing | Deep Learning | Extraction, Transformation & Load (ETL) | Machine Learning | Predictive Modeling | Data Visualisation Tools (Tableau, Qlikview, Hue) | Cloud Database (AWS, Azure)

LANGUAGES

English — Fluent

Mandarin — Conversational

OTHER SKILLS & CERTIFICATIONS

Skills

- Python Programming Language
- Machine Learning & Deep Learning by (Tensorflow, Scikit-Learn, PyMC)
- GCP Vertex AI with Kubeflow Pipelines
- BigQuery/Snowflake/PostgreSQL/AWS Athena/AWS Redshift
- Apache Airflow
- DBT
- Scrapy-Redis
- GCP Cloud Function / Cloud Run
- GCP Pub/Sub
- GCP Log Router
- GCP Looker
- Tableau
- Apache Superset
- Streamlit

Certification

- **Certified Data Science For Programmer,** issued by PT. Sharing Vision.
- **Google Cloud Big Data & Machine Learning Fundamentals**, issued by Coursera. Credential ID : JB4BK3H27Z6C
- **Getting Started with CyberGIS,** issued by Coursera**.** Credential ID : GN5P23GW3C5N
- **Dataiku ML Practitioner**, issued by Dataiku. Credential ID : u424bjg82q4o