

# Thimira Mastiyage Don

## Middle Data Scientist

### SUMMARY

- Mechatronics Engineer and Data Scientist with 4+ years in engineering data analytics, predictive modeling, and industrial operations.
- Expertise in Python, R, SQL, AWS, LangChain, Pinecone, and Power BI for scalable ML and data pipeline solutions.
- Delivered 70% reduction in manual reporting via automated ETL and real-time construction monitoring platforms.
- Developed RAG-based retrieval assistants and XGBoost risk prediction systems with 0.98+ recall deployed on AWS Lambda.
- Strong foundation in sensor/SCADA telemetry, time-series analysis, and engineering design using SolidWorks and MATLAB.

### TECHNICAL SKILLS

|                                                     |                                                                                                                                                                                                                                           |
|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Main Technical Skills</b>                        | Python (4 yr.), R (4 yr.), SQL (4 yr.), Scikit-learn (4 yr.), Xgboost (2 yr.)                                                                                                                                                             |
| <b>Programming Languages</b>                        | Python (4 yr.), R (4 yr.)                                                                                                                                                                                                                 |
| <b>AI &amp; Machine Learning</b>                    | AWS SageMaker (1 yr.), Hugging Face (1 yr.), LangChain (1 yr.), Mlflow (1 yr.), Neural Networks (1 yr.), NLP (3 yr.), NumPy (4 yr.), Pinecone (1 yr.), Prompt Engineering (1 yr.), PyTorch (1 yr.), Scikit-learn (4 yr.), Xgboost (2 yr.) |
| <b>Python Frameworks</b>                            | FastAPI (1 yr.)                                                                                                                                                                                                                           |
| <b>Python Libraries and Tools</b>                   | Matplotlib (3 yr.), NumPy (4 yr.), Pandas (4 yr.), PyTorch (1 yr.), Scikit-learn (4 yr.), Seaborn (3 yr.)                                                                                                                                 |
| <b>Mobile Frameworks and Libraries</b>              | Push notifications (4 yr.)                                                                                                                                                                                                                |
| <b>Data Analysis and Visualization Technologies</b> | DVC (1 yr.), Pandas (4 yr.), Power BI (3 yr.), Regression (4 yr.), Time Series (4 yr.)                                                                                                                                                    |
| <b>Databases &amp; Management Systems / ORM</b>     | MongoDB (2 yr.), PostgreSQL (3 yr.), SQL (4 yr.)                                                                                                                                                                                          |
| <b>Amazon Web Services</b>                          | AWS EC2 (1 yr.), AWS Lambda (1 yr.), AWS S3 (1 yr.), AWS SageMaker (1 yr.)                                                                                                                                                                |
| <b>Platforms</b>                                    | Arduino (1 yr.), Raspberry PI (1 yr.)                                                                                                                                                                                                     |
| <b>Deployment, CI/CD &amp; Administration</b>       | CD DevOps pipelines (1 yr.), CI/CD (1 yr.)                                                                                                                                                                                                |

|                                                     |                                                                               |
|-----------------------------------------------------|-------------------------------------------------------------------------------|
| <b>Virtualization, Containers and Orchestration</b> | Docker (1 yr.)                                                                |
| <b>SDK / API and Integrations</b>                   | FastAPI (1 yr.)                                                               |
| <b>Version Control</b>                              | Git (4 yr.), Github Actions (1 yr.)                                           |
| <b>Third Party Tools / IDEs / SDK / Services</b>    | MatLab (1 yr.)                                                                |
| <b>Other Technical Skills</b>                       | NX (1 yr.), PID controller (1 yr.), Solidworks (1 yr.), Unreal Engine (4 yr.) |

## WORK EXPERIENCE

### Data Scientist - RedBrick Engineering (Construction Monitoring and Material Passport Platform)

**Duration:** Sep 2023 – Nov 2025

**Summary:** Development of a real-time construction monitoring platform integrating structured and unstructured engineering/BIM documentation, and a Material Passport platform for capturing and retrieving material specifications and sustainability attributes across construction projects.

#### Responsibilities:

- Designed IFC data architecture and automated ETL data pipelines integrating engineering/BIM documentation.
- Built document and technical-data extraction from complex engineering models and manufacturer specification documents.
- Initiated a RAG-based retrieval assistant using Pinecone, LangChain, and NLP for project specifications.
- Contributed to data architecture of Material Passport platform.
- Developed interactive Power BI dashboards tracking real-time KPIs on progress and equipment/operational performance.

**Technologies:** IFC, ETL, Pinecone, LangChain, NLP, Power BI, BIM

### Manufacturing Technician - Illusion Australia Manufacturing Pty Ltd (Gas Log Fire Units Production)

**Duration:** Nov 2024 – May 2025

**Summary:** Applied engineering principles to equipment and process performance for the production of gas log fire units, ensuring safety compliance and quality assurance.

#### Responsibilities:

- Full unit assembly and production-equipment operation.
- Conducted gas leak and pressure testing on finished products.
- Ensured safety compliance and quality control throughout production.

### Data Scientist - MedHealth (IME Reports Data Conversion and Analysis)

**Duration:** Mar 2021 – Aug 2021



**Summary:** Converted unstructured IME reports into governance-compliant, analysis-ready datasets and standardized job titles for improved reporting consistency in healthcare data analytics.

**Responsibilities:**

- Developed custom Python parsers for document-based data extraction at scale.
- Standardised job titles into structured categories using TF-IDF clustering, MiniBatchKMeans, and PCA.
- Performed NLP extraction of Whole Person Impairment severity scores from clinical free-text.
- Conducted audit-style QA on data-driven outputs.
- Delivered Power BI dashboards for monthly stakeholder reviews.

**Technologies:** Python, TF-IDF, MiniBatchKMeans, PCA, NLP, Power BI

## **Data / Engineering Analyst - Dynamic Energies (Hydro Power Plant Risk and Performance Analysis)**

**Duration:** Jun 2018 – May 2019

**Summary:** Conducted engineering analysis and automated risk-calculation reporting for emergency events and system checks at a hydro power plant to support reliability and predictive maintenance.

**Responsibilities:**

- Analysed PLC and SCADA sensor/telemetry data, grid balance, transformer performance, and machine security.
- Applied statistical and time-series methods to operational and industrial data.
- Automated risk-calculation reporting for emergency events and system checks.

**Technologies:** PLC, SCADA, Statistical Analysis, Time-series Analysis

## **Mechatronics Intern - Zhejiang Jinlun Electro-Mechanic (Turbine Component Modelling and Process Analysis)**

**Duration:** May 2017 – Aug 2017

**Summary:** Modelled turbine components and analysed manufacturing and operational processes in hydro power stations to optimise equipment performance and quality assurance.

**Responsibilities:**

- Modelled turbine components using SolidWorks.
- Analysed processes in small and medium hydro power stations.
- Inspected turbine manufacturing processes and SCADA systems.
- Performed quality-assurance checks on turbine components.

**Technologies:** SolidWorks, SCADA

## **Mechatronics Intern - Wijeya Newspapers (Newspaper Production Process Improvement)**

**Duration:** May 2016 – Aug 2016



**Summary:** Developed solutions to resolve production issues and provided maintenance for newspaper printing machines to improve operational efficiency.

**Responsibilities:**

- Developed a solution to resolve auto-paste issues in newspaper production.
- Provided maintenance and servicing for printing machines.

## **EDUCATION**

- **MBA in Digital Transformation & AI, Royal Roads University, UAE**  
In progress  
2025 – 2027
- **M.Sc. Data Science, RMIT University, Australia**  
GPA 3.7 / 4.0  
2020 – 2021
- **B.Sc. Mechatronics Engineering, Asian Institute of Technology (AIT), Thailand**  
GPA 3.48 / 4.0  
2014 – 2018

## **CERTIFICATION**

- **AWS Certified Cloud Practitioner**
- **Google Data Analytics Professional**
- **IBM Databases & SQL**
- **Python for Data Science (IBM)**

