

M. Thufail Alwannabil Samas

Data and AI Professional | Machine Learning, Data Systems and Applied AI

Surabaya, East Java, Indonesia | mthufailalwannabils@gmail.com | +62 877-6313-3209

mthufailsamas.github.io | linkedin.com/in/mthufailsamas | github.com/mthufailsamas

SUMMARY

Data and AI professional with experience building Python/PostgreSQL data pipelines, machine-learning and time-series systems, and applied AI automation workflows. Built solutions for optical sensor monitoring, weather forecasting, classification, and LLM-assisted invoice processing with API integrations, database-backed workflows, and human review. BNSP-certified Data Scientist and first author of a peer-reviewed XGBoost rainfall forecasting publication.

EXPERIENCE

PT Aplikasi Lintasarta

Junior Programmer

Surabaya, Indonesia

January 2026 - April 2026

- Developed a project-stage DWDM optical monitoring and assurance system for approximately 400 directional sensors, integrating five-minute PRTG telemetry with ENIMS network metadata to identify operational failures, persistent signal shifts, gradual degradation, and investigation priorities.
- Built an incremental Python and PostgreSQL pipeline to synchronize metadata, collect telemetry, repair data gaps, generate processed observations, detect events, compare paired sensors, rank sensor priorities, and serve dashboard data through an API.
- Implemented data-quality and failure-detection rules for missing readings, low optical power, downtime, and zero coverage, grouping adjacent five-minute failure intervals into continuous events and excluding invalid observations from statistical analysis.
- Developed sensor-adaptive analytics using robust six-hour medians, PELT change-point detection, Theil-Sen slope estimation, and Mann-Kendall trend testing to distinguish persistent signal shifts and gradual degradation from short-term telemetry noise.
- Created a browser-based monitoring interface with Operations Board, Event Explorer, and Management Overview views, allowing users to review active failures, statistical events, paired-sensor evidence, investigation charts, and current sensor priorities.

SMP GIKI 3 Surabaya

Coding and AI Teacher

Surabaya, Indonesia

August 2025 - December 2025

- Taught Python fundamentals to Grade 8 and 9 students, covering variables, data types, operators, conditionals, loops, and programming logic through structured lessons and hands-on coding exercises.
- Introduced foundational AI concepts using Teachable Machine, guiding students through datasets, model training, and image-classification experiments to demonstrate how machine-learning models work in practice.
- Developed lesson materials, coding exercises, quizzes, and sample programs, using practical assignments and structured assessments to support instruction and evaluate student understanding.

Universitas Islam Negeri Sunan Ampel Surabaya (UINSA)

Programming Teaching Assistant

Surabaya, Indonesia

August 2022 - July 2024

- Guided undergraduate students across four semesters of programming courses in MATLAB, Python, and computational problem-solving through practical exercises and assignment support.
- Supported students in debugging, data handling, and algorithmic thinking, helping them turn mathematical and computational concepts into working code for exercises and assignments.
- Developed introductory machine-learning materials covering Linear Regression, K-Nearest Neighbors, and K-Means clustering, and provided structured technical feedback on student assignments.

Stasiun Meteorologi Maritim Kelas II Tanjung Perak Surabaya (BMKG)

Data Science Intern

Surabaya, Indonesia

February 2024 - June 2024

- Developed a one-page extreme-weather early-warning dashboard using HTML, CSS, JavaScript, PHP, MySQL, and periodic API updates for BMKG's forecasting workflow.
- Processed and compared BMKG cloud-prediction imagery with Himawari IR satellite imagery using Python, supporting descriptive assessment of cloud movement and temperature patterns.
- Built a three-layer BiLSTM model for next-day rainfall forecasting using seven-day multivariate sequences, a focused grid search, and an 80:20 chronological split, achieving MAAPE of 0.81 and RMSE of 10.27 mm.

EDUCATION

Universitas Islam Negeri Sunan Ampel Surabaya (UINSA)

Surabaya, Indonesia

Bachelor of Mathematics (S.Mat.)

August 2021 - January 2025

- **GPA:** 3.46/4.00
- **Undergraduate Thesis:** Rainfall Forecasting in Sumenep Regency Using XGBoost and Grid Search.
- **Academic Highlight:** Completed the 4-year bachelor's program in 7 semesters (3.5 years).
- **Relevant Coursework:** Artificial Intelligence, Programming, Mathematical Statistics, Statistical Methods, Multivariate Analysis, Numerical Methods, and Applied Linear Algebra.

SKILLS

- **Programming & Query:** Python, SQL, R, MATLAB, PHP
- **Data Processing & Databases:** Pandas, NumPy, PostgreSQL, MySQL, psycpg, Data Ingestion, Data Validation, Data Transformation, Data Pipelines
- **Machine Learning & Statistics:** scikit-learn, XGBoost, TensorFlow/Keras, statsmodels, Statistical Analysis, Feature Engineering, Model Evaluation, Time-Series Forecasting
- **Applied AI & Automation:** Gemini API, n8n, FastAPI, Pydantic, REST APIs, LLM-Assisted Document Processing, Workflow Automation, Human Review Workflows
- **Development & Visualization:** Docker Compose, Git, GitHub, Jupyter Notebook, unittest, Matplotlib, Seaborn, HTML, CSS, JavaScript

PROJECTS

AI Invoice Approval Automation

Built a batch invoice-processing and approval workflow with Gemini, n8n, FastAPI, PostgreSQL, and Google Sheets. It processes up to 20 PDF or image files, validates extracted data, prevents duplicates, routes exceptions for review, records an audit trail, and sends approved data to an ERP sandbox.

F&B Demand Forecasting

Built a reproducible weekly demand-forecasting pipeline for 456,548 observations across 145 weeks, 77 fulfilment centers, and 51 meals. Compared a four-week baseline with XGBoost using chronological evaluation, reducing final-test MAE by 32.6%, from 113.53 to 76.55 orders.

Thyroid Cancer Recurrence Classification

Built an XGBoost classification workflow for thyroid cancer recurrence using 383 UCI records, mixed-type preprocessing, Information Gain, grid search, and 10-fold cross-validation. Achieved pooled out-of-fold performance of 97.25% accuracy, 95.37% F1-score, and 98.60% ROC-AUC.

PUBLICATIONS

Prediksi Curah Hujan di Kabupaten Sumenep Menggunakan Metode *Extreme Gradient Boosting* (XGBoost) dan Algoritma *Grid Search*

First Author | MIND Journal, Vol. 11, No. 1, pp. 30-43 | June 2026 | DOI: 10.26760/mindjournal.v11i1.30-43

CERTIFICATIONS

Data Scientist - Badan Nasional Sertifikasi Profesi (BNSP) | September 2024 - September 2027