Samuel Mwania

Software Engineer · Machine Learning and Data Systems

mwaaniasamuel@gmail.com · +254 706 203 624
mwaaniasam.com · github.com/mwaniasam
linkedin.com/in/mwaaniasam · Juja (Nairobi), Kenya

Completing December 2026. Open to internships and project work now and full-time roles from January 2027; remote or relocating.

SUMMARY

Software engineer who takes machine learning and data problems from raw data to a running service that is evaluated against a stated baseline, deployed, tested and documented. Placed 17th of 1,651 on Zindi, runs two ML services in production, and built a procurement-integrity capstone with pre-registered results on Kenyan and Estonian data.

EDUCATION

BSc Software Engineering, Machine Learning specialisation · African Leadership University — Jan 2024 – Dec 2026

Capstone: AegisVeritas, finding hidden supplier links in Kenyan public procurement. Supervisor: Kevin Sebineza.

EXPERIENCE

Data Engineer Intern · Hello Tractor, Nairobi — Apr 2025 – Aug 2025

Agritech platform connecting smallholder farmers with tractor owners across Africa.

- Built and prepared structured datasets for internal analysis and downstream reporting.
- Audited source data for missing fields, inconsistencies and format errors before load, keeping bad records out of reports.
- Supported batch prediction runs, organising model inputs and checking outputs against expected formats, and documented the procedure for engineering-to-analytics handover.

SELECTED PROJECTS

AegisVeritas · capstone · mwaaniasam.com/aegisveritas — 2026

Finds bidders in Kenyan public procurement that look independent but share an owner, or link to a debarred party.

- Scores supplier pairs from shared owners and directors, shared contacts and award behaviour (calibrated LightGBM), combined Fellegi-Sunter style so every flag keeps its reasons; 12,441 disclosed ownership links as labels.
- Pre-registered tests: average precision **0.146** against 0.067 for the Adamic-Adar baseline on 7,884 held-back Kenyan pairs, with 25 of the top 100 flags real links; AUC **0.85** against 0.57 on Estonian firms the model never saw.
- FastAPI, PostgreSQL with row-level security, Neo4j, Celery, Redis, Next.js and a hash-chained audit log; **665** backend tests, 80 of 80 browser tests. Expert evaluation under way.

Multilingual Health Question Answering · Zindi competition — **17th of 1,651** · 2026

- Health questions in Akan, Amharic, Luganda, Swahili and English: eight language–country subsets, 36,501 curated pairs.
- Dense retrieval with a fine-tuned AfriE5-Large embedder and FAISS, then per-language answer selection: consensus voting, cross-encoder reranking, or Qwen2.5-7B LoRA generation where it scored better.
- Public score from 0.4908 to **0.6908** (+41%) over 44 logged experiments, 24 rejected and kept in the log; open-source only.

CoffeeGuard · coffee leaf disease classifier · live — 2026

- MobileNetV2 transfer learning on 58,549 images in five classes: **98.80%** accuracy and 98.36% macro F1, so the rare leaf rust class cannot hide behind the headline.
- Eight FastAPI endpoints, a Streamlit dashboard and retraining history in SQLite, on Docker Compose and Render; zero failed requests in Locust load tests at 10 and 50 concurrent users.

Maize Yield Prediction · API and Flutter app · live — 2025

- Random Forest served through six validated FastAPI endpoints to a Flutter app, trained on 17,516 records from 37 Nigerian states. R^2 of 0.24 reported as is, with the cause (no rainfall or soil features) and the fix documented.

Milan Mobile Traffic Forecasting — 2026

- Streamed 5 GB of telecom telemetry from 10,000 grid squares into a 28 MB Parquet file under 3 GB of memory. On the benchmark area an LSTM cut error from SARIMA's 43% MAPE to 7.5%; a TCN trained three times faster on GPU.

Also: Freeway DQN (team of three; my 10 of 30 RL experiments showed discount 0.90 beating 0.99) · Synapse, adaptive learning in Next.js and PostgreSQL · MoMo Analyst (team of four; I built the SMS-to-MySQL data pipeline).

SKILLS

Machine learning	PyTorch, TensorFlow and Keras, scikit-learn, LightGBM, Hugging Face Transformers, PEFT, sentence-transformers, FAISS, Stable Baselines3
Data and backend	Python, SQL, PostgreSQL, Neo4j, MySQL, pandas, Parquet, FastAPI, Flask, Celery, Redis
Engineering	Docker, CUDA, Linux, Git, pytest, Playwright, Locust, TypeScript, Next.js, Flutter