Calistus Mwonga

Data Scientist

Athi-River, Machakos County | +254706220356 | calistusmwonga@gmail.com |

LinkedIn | GitHub | Portfolio

SUMMARY

Data Scientist with a Computer Science foundation and expertise across the full data lifecycle: data analysis, machine learning and model deployment. Experienced in building product-ready solutions that solve real-world business problems.

SKILLS

Languages: Python, SQL (MySQL, PostgreSQL)
Data analysis & Visualization: Pandas, NumPy, SciPy, Statsmodels, Matplotlib, Seaborn, Plotly, Tableau, Power BI
Machine Learning: Scikit-Learn, TensorFlow, Keras. Pytorch, SHAP, Lime, Optuna
Deployment & Cloud: FastAPI, Streamlit, Docker, REST APIs, Linux, Git

PROJECTS

Landslide Classification System | Python, FastAPI, TensorFlow, SHAP, Streamlit, LightGBM

View project

- Built a computer vision system (F1-Score: 0.86) to detect landslides by combining optical (Sentinel-2) and all-weather radar (Sentinel-1) satellite data, solving the critical problem of detecting during cloud cover.
- Compared a deep learning (CNN) approach against a traditional ML model (LightGBM), demonstrating a robust and data-driven model selection process.
- Deployed the best model as a full-stack application with a FastAPI backend and a Streamlit frontend that delivers real-time, per-prediction model explanations using SHAP, turning a black box into a transparent tool.

AI First-Aid Assistant | Python, FastAPI, RAG, Next.js, TypeScript, PostgreSQL, PostGIS

View project

- Built a full-stack AI assistant using a Retrieval-Augmented Generation (RAG) pipeline to provide accurate, context-aware first-aid guidance from a verified knowledge base.
- Engineered a multi-stage Responsible AI safety pipeline that first validates user intent and then analyzes the AI's own output to automatically detect and escalate high-urgency medical emergencies.

- Deployed the system with a modern web stack, featuring a Next.js/TypeScript frontend and a Python/FastAPI backend with a PostGIS-enabled geospatial database for an integrated 'Find Hospital' feature.

Loan Default Prediction & Credit Scoring System | Python, Scikit-learn, XGBoost, Pandas

View project

- Developed and tuned a best performing model (XGBoost) to predict loan defaults, achieving a strong F1-Score of 0.75 on a highly imbalanced dataset with a default rate of less than 2%.
- Engineered better predictive features by performing customer-level aggregation to create historical behavior profiles for each borrower (e.g., average loan size, loan count).
- Translated raw model probabilities into an actionable business tool by designing and implementing a 5-tier credit scoring system to automate risk assessment and guide lending decisions.

EDUCATION

Moringa School | **Nairobi, Kenya**
Data Science Certificate | *February 2025 - August 2025*

Dedan Kimathi University of Science and Technology | **Nyeri, Kenya**
Bachelor of Science in Computer Science | November 2020 - *November 2024*

St. Thomas Moore Nguviu Boys High School | Embu, Kenya

Secondary Level of Education | February 2015 – November 2018

VOLUNTEERING

ST. James Ngimu Church Ministry | Athi-River Kenya

Technician May 2024 – Present

- Setup and display multimedia content including songs and presentations for Sunday services
- Ensure timing and smooth transitions so the congregation can easily follow the service
- Troubleshoot technical issues beforehand and maintain a stable connection to ensure a smooth service flow