



CLEAVESTONE ADUNGO

cleavestone94@gmail.com | +254703457427 | Nairobi Kenya

GitHub: [Link](#) | LinkedIn: [Link](#) | Portfolio: [Link](#)

Summary

Data Scientist transitioning from finance with 2+ years of hands-on ML, NLP, and data analytics experience. Skilled in Python, SQL, statistical modeling, LLM fine-tuning, RAG systems, and MLOps pipelines. Diverse industry background (education, logistics, finance, data annotation) with a proven track record of turning raw data into actionable insights and delivering production-ready ML solutions. Competitive ML participant (top 30% Zindi), passionate about solving high-impact problems with data.

Skills

Programming & Analytics: Python (pandas, NumPy, scikit-learn, matplotlib, seaborn), SQL, Excel

Machine Learning & AI: Classification, Regression, NLP, Deep Learning, Feature Engineering, Model Evaluation

LLMs & Conversational AI: LangChain, RAG, Vector Databases (Pinecone, LanceDB), Prompt Engineering

Data Engineering & BI: ETL (Airflow), MongoDB, Power BI, Streamlit, Data Visualization

MLOps & Deployment: MLflow, DVC, Docker, Git/GitHub, AWS (EC2, S3, RDS, ECR), CI/CD, Model Monitoring

Soft Skills: Analytical Thinking, Problem-Solving, Communication, Self-Directed Learning

Experience

Zindi Africa – Data Science Competitions *April 2025 – Present*

- Participated in multiple machine learning competitions tackling real-world problems in healthcare and agriculture.
- Applied methods such as classification, regression, feature engineering, and LLM fine-tuning (LoRA, QLoRA).
- Gained hands-on experience handling data drift and distribution shift, improving model robustness through cross-validation
- and by preventing data leakage.
- Awarded a performance voucher in the *Amini Soil Challenge* for developing a model to **predict** macro and micro nutrients in the soil.

AI Data Trainer | Turing (Remote, USA) | *Oct 2024 – Apr 2025*

- Annotated large-scale datasets to train ML models, improving data quality and model accuracy.
- Designed & tested **LLM prompts** to boost clarity, relevance, and precision of responses.
- Ensured high data integrity through QA checks and efficient use of annotation tools.

Accounts Analyst (Payables & Receivables) | Signature Healthcare Ltd (Kenya) | *Jun 2021 – Oct 2024*

- Automated financial dashboards (Excel/Power BI), improving reporting efficiency by 50%.
- Built **payment lookup system in Excel (VLOOKUP)** eliminating manual entry errors.
- Analyzed operational datasets (mileage, service logs, spare parts), enabling data-driven maintenance decisions.
- Developed KPIs for tracking receivables and collections, improving debt recovery timelines.

Education and Training

Moi University – Eldoret, Kenya

Bachelor of Science in Mathematics | 2012 – 2017

Second Class Honors, Upper Division

Relevant coursework: Probability & Statistics, Linear Algebra, Real Analysis, Numerical Methods, Mathematical Modeling, Calculus

Projects

Customer Churn Prediction – End-to-End MLOps Pipeline

Built a production-ready churn prediction system for a banking dataset using LightGBM, achieving an ROC AUC of 0.86. Designed a modular ML pipeline (data ingestion → preprocessing → training → evaluation) with MLflow for experiment tracking and DVC for data versioning. Containerized the application with Docker and integrated CI/CD (GitHub Actions → AWS ECR/ECS Fargate) for scalable deployment. Implemented a FastAPI backend with a web UI and Grafana for drift monitoring [Link](#)

Movie recommender app - a semantic movie recommendation system powered by LanceDB, Sentence Transformers, and Streamlit. This app suggests movies based on the similarity of user queries to movie descriptions using vector embeddings. [Link](#)

Amini Soil Challenge hosted on the Zindi platform, ranked 40th out of 400 participants in a machine learning competition focused on predicting soil nutrient availability and crop yield optimization.

ETL Pipeline: This project builds a reliable ETL pipeline using Apache Airflow and Python. It extracts real-time weather data from the OpenWeatherMap API, transforms it into a structured format, and loads it into MongoDB Atlas. The setup enables efficient storage and future analysis of weather trends in the cloud. [Link](#)

Data Analysis (SQL) - Built an interactive SQL practice project with realistic datasets to strengthen querying skills. It features three tables: customers, products, and orders. The project covers essential SQL topics like JOINS, GROUP BY, subqueries, and window functions through practical exercises. [Link](#)

Crop Type Classification with Satellite Imagery - Ranked 26th out of 170 in a Zindi challenge to classify crops in Côte d'Ivoire using Sentinel-2 satellite imagery. Built a time-aware model to distinguish between cocoa, rubber, and oil palm. Contributed to scalable, remote sensing solutions for sustainable agriculture.

Modelling Clinician Response in Rural Kenya - Ranked 70th out of 450 in a Zindi NLP competition simulating clinician decision-making in rural Kenyan hospitals. Fine-tuned a FLAN-T5 model on expert-annotated clinical scenarios using prompt engineering and domain-specific preprocessing.

Northwind Sales Analysis – Power BI Dashboard - Designed interactive Power BI dashboard analyzing Northwind sales database on SQL Server. Performed data cleaning/transformation with Power Query, built relational data model with DAX measures, and created KPIs for sales, profit, and delivery metrics. Published to Power BI Service with automated refresh for stakeholder access. [Link](#)

Certifications

- Certificate in MLOPS from Udemy
- Certificate in LLM Engineering from Udemy
- Certificate in Foundational Generative AI from Ineuron