

JOHN APOLLOS OLAL ONYANGO

Data Scientist & AI/ML Engineer

Nairobi, Kenya | +254 745 648512 | johnapollosolal@gmail.com | apollos-digital.vercel.app
linkedin.com/in/johnapollos | github.com/JohnApollos | medium.com/@johnapollosolal

PROFESSIONAL SUMMARY

Final-year BSc Data Science and Analytics student at JKUAT with hands-on experience building end-to-end ML pipelines, deploying production APIs, and engineering full-stack data-driven applications. Proficient in Python, scikit-learn, FastAPI, Docker, and CI/CD workflows. Track record of translating business problems into deployed ML solutions across fintech, health, and enterprise software domains. Actively developing MLOps capabilities including experiment tracking with MLflow and model monitoring. Seeking a junior AI engineering role where applied machine learning, software development, and cross-functional collaboration converge.

CORE TECHNICAL COMPETENCIES

Data Science & ML:	Python (pandas, NumPy, scikit-learn), Supervised/Unsupervised Learning, MLflow experiment tracking (in progress), hyperparameter tuning (GridSearchCV), model evaluation (McNemar's test, F1, ROC-AUC), SHAP explainability, XGBoost, Regression, Time Series, Evidently drift monitoring.
Data Engineering & Quality:	SQL (PostgreSQL, SQLite), NoSQL (MongoDB), custom ETL/ELT pipelines, unstructured data parsing (Regex, OCR), data cleaning & validation (Pydantic schemas), ReportLab PDF generation.
Geospatial & Visualization:	GeoPandas, Folium, Leaflet, choropleth mapping, spatial analysis, Plotly, Matplotlib, Seaborn, Recharts, Power BI, Streamlit.
Mobile & Frontend Development:	Flutter, Dart, offline-first architectures, Next.js, React.js, TypeScript, JavaScript, Tailwind CSS, Dexie.js (IndexedDB).
Backend & Infrastructure:	FastAPI, Node.js, REST APIs, Socket.IO, Supabase, JWT auth, Row-Level Security (RLS), Git, GitHub Actions CI/CD, Docker containerisation, Linux/Debian, Railway, Vercel, Render.

PROFESSIONAL EXPERIENCE

Founder and Lead Engineer

Oct 2025 - Present

Apollos Digital Solutions | Nairobi, Kenya

- Built and deployed full-stack data products, SaaS platforms, AI automation tools, and ETL pipelines for clients across financial services, civic intelligence, legal, travel, and automotive sectors.
- Usler (Fintech ETL):** Engineered Python/Regex ETL pipelines parsing highly unstructured M-Pesa SMS transaction strings into clean database entries. Structured data in PostgreSQL and built Power BI analytical dashboards.
- GaragePulse (SaaS):** Deployed an edge-based OCR AI plate reader for workshop vehicle capture. Integrated strict Pydantic models for data validation to prevent corrupted records.
- Multi-Tenant SaaS:** Configured PostgreSQL schemas with Row-Level Security (RLS) policies and JWT authentication to enforce strict tenant isolation across shared database instances.
- SupaKonekt (Marketplace):** Built a service platform integrating Safaricom Daraja M-Pesa API STK pushes, Socket.IO real-time chat, Nodemailer notification services, and automated cron subscription billing cycles.
- AW Client Portal (Fintech):** Created an automated quarterly PDF report generation tool with Python FastAPI and ReportLab, reducing financial advisor compilation time from 8 hours per report to 5 seconds.

Civic Data Consultant

Early 2026 - Present

Marsabit and Nyakach Strategy Initiatives | Kenya

- Built automated Python ETL pipelines extracting and cleaning IEBC voter registration data across 446 polling stations for two 2027 electoral campaigns.
- Performed geospatial analysis and demographic mapping using GeoPandas and Folium to highlight swing wards.
- Produced 31-page intelligence report maps and Excel workbooks for campaign leadership.

Deployment Coordinator & Web Developer

Jan 2026 - Mar 2026

Lakwet Ap Rose World Travellers | Remote

- Managed live deployments and CI/CD pipelines using GitHub Actions for a commercial travel monorepo.
- Enforced version control and code quality standards via strict Git workflows.

KEY PROJECTS & CASE STUDIES

FinLens AI (FastAPI, Gemini, LangChain)

Financial risk analysis engine extracting metrics via Gemini, validating balance sheets, and running local RAG anomaly models logged to MLflow.

SupaKonekt Marketplace (React, Node.js)

Service marketplace with M-Pesa STK push payments, real-time Socket.IO chat, and subscription management.

NyaPulse Campaign Intel (React, Zustand)

Nyakach 2027 MP campaign dashboard with competitor SWOT grid, scenario simulators, and targeting maps.

LexPulse Practice SaaS (Flutter, FastAPI)

Advocate practice manager compliant with Kenya Data Protection Act. Matter billing and AI brief summaries.

Customer Churn Predictor (XGBoost, MLflow)

Fintech churn prediction model with SHAP explainability, Evidently drift monitoring, and Streamlit deployment.

Kenyan Healthcare Analysis (GeoPandas)

Mapped facility density and ownership dependencies across 47 counties using Folium choropleth layers.

Kenya Boda Boda E-Mobility Viability Engine (GeoPandas, scikit-learn)

- Independent Research Project, July 2026
- Built a five-component data framework for battery swap network expansion, PAYG default risk modelling, and optimal station placement across Nairobi's electric motorcycle sector.
 - Coverage gap analysis across 20 swap stations using GeoPandas buffer zone analysis, identifying 279 underserved grid points.
 - Logistic regression PAYG default model showing distance to nearest swap station is 2.17x stronger default predictor than income level.
 - K-Means optimal placement model reducing simulated portfolio default risk by 14% and average swap travel distance by 1.8km.

Fertilizer Dispatch Analysis (Plotly)

NCPB 2025 Long Rains agricultural records cleaned and visualized with Plotly county dashboards.

FirstAid+ Offline App (Flutter, SQLite)

Offline capstone mobile platform with optimized local SQLite database for zero-latency queries.

AW Client Report Portal (FastAPI, ReportLab)

Automated cashflow (SACS) & net worth (TCC) report generation form-to-PDF pipeline for a US wealth firm.

BaseOps Fleet Logistics (Next.js, Supabase)

Logistics PWA featuring tenant isolation via Supabase RLS and Dexie.js offline-first mutation queues.

GaragePulse SaaS (FastAPI, PostgreSQL)

Repair shop manager with edge OCR number plate scanning, digital Gate Pass, and WhatsApp quote generation.

Hustler OS / Usler (Flutter, Python)

ETL regex statement parser pipeline converting raw M-Pesa SMS strings into PostgreSQL database entries.

Toxicity ML Pipeline (Python, scikit-learn)

High-dimensional chemical classification pipeline using L1 regularization, SMOTE, and GridSearchCV.

UCI Seeds Analysis (scikit-learn)

Unsupervised K-Means/Hierarchical clustering and supervised classifiers compared via McNemar's testing.

Credit Scoring Predictor (scikit-learn)

Logistic Regression loan default predictor achieving 79% accuracy and 95% default detection precision.

Startup Investment EDA (scikit-learn)

Unsupervised K-Means clustering and PCA analysis on startup funding and investment trajectories.

EDUCATION & CREDENTIALS

BSc in Data Science and Analytics

JKUAT Karen Campus | Expected Nov 2027
Class Representative | GPA: In Progress
ML I & II, AI, Time Series, Stats Inference

Software Engineering Certificate

Power Learn Project Africa | Completed 2025
Python, JS, SQL, React, Dart, Flutter

Virtual Assistant Certification

ALX Africa | Completed Nov 2024
Remote operations, task tracking, agile workflows

ADDITIONAL INFORMATION

Hackathons & Awards:

Presented Hustler OS at US Embassy Nairobi Red White & Build Hackathon (Jan 2026)

Interests & Focus:

MLOps & production ML, fintech/insurtech
AI, geospatial engineering

Languages & References:

English (C2), Kiswahili (C1), Dholuo (Native).
References available on request.