

Data Scientist specializing in production NLP systems, real-time recommendations, and scalable data infrastructure. apt at building end-to-end ML pipelines, from model development through deployment and monitoring, across consumer apps, healthcare analytics, and enterprise platforms.

Work Experience

Data Scientist Data Team	Chub Remote	May 2024–Dec 2025
<ul style="list-style-type: none">Deployed NLP models for community content moderation and sentiment analysis using transformer architectures, improving detection accuracy by 24% and reducing manual review overhead by 35% across user-generated posts and challenge comments.Engineered FastAPI microservices with lazy loading, streaming, and caching strategies to serve real-time ML predictions for habit recommendations, reducing P95 latency by 45% under peak loads.Integrated Elasticsearch for challenge discovery and user search, optimizing query performance from 1.2s to <250ms through strategic indexing and relevance tuning.Built CI/CD pipelines with GitLab, Docker, and Terraform for automated model training and deployment on AWS, reducing release cycles from weekly to daily with 99.9% uptime via MLflow and CloudWatch monitoring.Designed ETL workflows using Airflow and AWS Batch to process daily user interactions, enabling engagement trend analysis and personalized recommendation features.		
Data Scientist Data Team	TD Africa, Lagos, Nigeria	May 2023–Apr 2024
<ul style="list-style-type: none">Automated ERP data extraction from Microsoft Dynamics 365 F&O, reducing reporting latency from 2 days to 1 hour.Designed star-schema data models in Power BI with incremental refresh, supporting over 1000 concurrent users and 50+ reports across sales, inventory, and procurement teams.Developed time-series forecasting models to predict stock levels, reducing inventory overstock by 18%.Orchestrated multiple ELT pipelines in Airflow with Docker containerization, implementing data quality checks and SLA monitoring that reduced pipeline failures by 4%.Analyzed data for the marketing team to evaluate campaign effectiveness and regional outreach strategies, contributing to a 30% increase in testing center footfall during key periods.Facilitated internal workshops to train 100+ staff in SQL, Power BI, and data-driven business practices.		
Data Analyst Data Team	Medbuzz Lagos, Nigeria	Feb 2021–Mar 2023
<ul style="list-style-type: none">Led the migration from Excel-only workflows to a PostgreSQL database (via Supabase), improving data reliability and enabling advanced SQL querying for operational and financial analysis.Built and maintained Power BI dashboards to provide daily insights on testing volumes, positivity rates, and turnaround times, empowering executives and operations teams with real-time data for rapid decisions during the COVID-19 peak.Created automated reports for state health authorities and internal stakeholders, reducing manual reporting time by 80% and ensuring timely compliance submissions.Developed forecasting models in Power BI to predict test kit demand based on historical trends and regional infection rates, helping procurement teams reduce stockouts and wastage.Provided ad-hoc analyses to support operational decision-making, including staff allocation optimization, center performance benchmarking, and customer wait time reduction.		

Education and Certifications

- **B.Sc., Metallurgical and Materials Engineering**, University of Lagos, Nigeria. **2014–2020**
- **Microsoft Certified: Fabric Analytics Engineer Associate**
- **DataCamp Certified: Data Analyst Associate**

Technical Skills

- **Programming:** Python, SQL, Bash.
- **Visualization:** Tableau, Power BI, Excel, Looker, Streamlit.
- **Orchestration & Cloud:** AWS (Bedrock, Batch, Lambda, S3), Docker, Terraform, Airflow
- **Data Science & ML:** Scikit-learn, Pandas, NumPy, NLP, SHAP, SMOTE, OCR, FASTAPI, OLLAMA, QDRANT.
- **DevOps:** Git, GitHub, Terraform.
- **Soft Skills:** Stakeholder Communication, Team Leadership, Analytical Thinking.

Projects

Fraud Prediction Deployment Github

Stack: Python, FastAPI, scikit-learn, SQLite

- Deployed an end-to-end fraud detection system, achieving 82% precision and 79% recall using ensemble methods (Random Forest + XGBoost), processing 1,000+ transactions daily with sub-200ms prediction latency through optimized feature engineering pipelines.
- Designed 25+ behavioral features, including rolling window aggregations, merchant risk scores, geolocation anomaly detection, transaction time patterns, and device fingerprinting to capture complex fraud patterns across multiple dimensions.
- Built production-grade REST API with FastAPI featuring async request handling, request validation with Pydantic, rate limiting, API key authentication, comprehensive error handling, and automated OpenAPI documentation for seamless integration with payment systems.
- Designed SQLite database schema with transaction history, prediction logs, and fraud labels, implementing efficient indexing strategies and automated data retention policies for compliance while maintaining query performance under growing data volumes.

MLOps Monitoring Pipeline Github

Stack: Python, FastAPI, MLflow, Evidently AI, Docker, Prometheus, scikit-learn

- Built MLOps pipeline demonstrating end-to-end ML workflow from training to deployment, featuring experiment tracking with MLflow, real-time prediction API serving 100+ requests/second, and automated drift detection generating HTML reports for model performance monitoring.
- Engineered FastAPI microservice with Prometheus metrics, health checks, and batch prediction support, containerized with Docker Compose, achieving 99% uptime through automated health checks, restart policies, and comprehensive OpenAPI documentation.
- Implemented data drift detection using Evidently AI to monitor model degradation, tracking distribution shifts across text features with automated alerting, and reducing manual model validation effort through continuous monitoring dashboards.