

EDIDIONG DAVID IBOKETE

DATA SCIENTIST | MACHINE LEARNING ENGINEER | GENERATIVE AI, GenAIOps & BUSINESS INTELLIGENCE

Houston, TX | U.S. Permanent Resident | No Sponsorship Required | iboketee@icloud.com

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

PROFESSIONAL SUMMARY

Data scientist and machine learning engineer with about three years of experience building Generative AI, machine learning, data engineering and business intelligence solutions across education, healthcare, research libraries and public-sector operations. Microsoft-certified across Azure MLOps, Power BI analytics and Azure cloud fundamentals; builds version-controlled Python and SQL systems, Power BI semantic models and Azure-supported GenAI workflows with CI/CD, model evaluation, cloud deployment and stakeholder reporting.

TECHNICAL SKILLS

Programming & Data Engineering: Python, SQL, R, SAS PROC SQL, Pandas, NumPy, Excel/VBA, Git, GitHub, REST APIs, Jupyter, ETL, data wrangling, feature engineering, data validation, reproducible pipelines

Machine Learning & Statistics: Scikit-learn, XGBoost, TensorFlow/Keras, regression, classification, forecasting, anomaly detection, clustering, NLP, model evaluation, SHAP, residual diagnostics

Generative AI, MLOps & Cloud: Generative AI, GenAIOps, RAG, LLMs, embeddings, AI agents, semantic retrieval, prompt engineering, FastAPI, Azure App Service, Azure Machine Learning, MLflow, Microsoft Foundry, model registration, online endpoints, experiment tracking, GitHub Actions CI/CD, Docker, OpenTelemetry, Application Insights

Business Intelligence & Platforms: Power BI, Power Query/M, DAX, CALCULATE, filter context, time intelligence, semantic modeling, star schema, Power BI Service, workspaces/apps, scheduled refresh, gateways, row-level security, Tableau, MS SQL Server, MySQL, SLATE CRM, Alteryx, ArcGIS

EDUCATION & CERTIFICATIONS

M.S., Data Analytics (Big Data), Canisius University, Buffalo, NY | Completed December 2025

B.S., Economics, University of Abuja, Nigeria

Microsoft Certified: Machine Learning Operations Engineer Associate (AI-300) | Power BI Data Analyst Associate (PL-300) | Azure Fundamentals (AZ-900)

PitchBook Pioneer: Core Foundations Certification | 2026

PROFESSIONAL EXPERIENCE

Canisius University | Buffalo, NY | Aug 2024 - Dec 2025

Graduate Assistant, Data Analytics & Reporting

- Partnered with admissions, enrollment and leadership stakeholders to identify, scope and prioritize data opportunities, translating operational questions and reporting pain points into analytical requirements, trusted datasets, KPIs and decision-support solutions.
- Built and maintained ETL, reconciliation and data-quality workflows across SLATE CRM and MS SQL Server using SQL, SAS PROC SQL, R, Alteryx and Excel, achieving 100% deduplication accuracy and reducing record redundancy by 70%.
- Designed relational Power BI semantic models with fact and dimension tables and star-schema relationships; used Power Query/M to profile, clean, merge, append and reshape data and DAX (CALCULATE, filter context and time intelligence) to create reusable KPI measures.
- Delivered interactive Power BI and Tableau dashboards with filters, slicers, drillthrough and tooltips for recurring and ad hoc reporting; documented data flows, KPI definitions and validation rules and communicated findings to technical and nontechnical stakeholders.

Center for Research Libraries | Chicago, IL | Jun 2025 - Aug 2025

Algorithm & Data Engineer Intern

- Partnered with stakeholders across 19 institutions and data engineers to define requirements and translate an ambiguous cooperative-cataloging problem into a Python/Pandas matching and allocation solution spanning 194 tasks and 565 candidate matches.
- Acquired, profiled, normalized, cleansed, transformed and validated multi-source survey data, resolving inconsistent schemas and building reliable requester and contributor data models for algorithm development and operational decision support.
- Designed, tested and tuned rule-based matching logic around capacity, qualifications, reciprocity and trust, then conducted exception analysis and documented assumptions, limitations and validation outcomes.
- Developed modular Python and SQL components under Git version control, with anonymized sample data, configuration, exports, validation checkpoints and an end-to-end smoke test for AWS-supported technical review and reproducible handoff.

Pear Project Inc. | Remote | May 2025 - Jul 2025

Applied AI & Healthcare Data Science Intern

- Co-developed an Azure-supported conversational Generative AI patient-intake and provider-routing prototype for women's reproductive health, translating unstructured symptom narratives into specialist recommendations across the platform's provider database.
- Engineered cloud-based healthcare data workflows in Azure using Python and R to acquire, cleanse, transform and validate unstructured symptom descriptions, structured provider records, data dictionaries and ICD-10, HCPCS and CPT reference artifacts.
- Developed and evaluated NLP, text-classification, predictive-modeling and prompt-engineering workflows, applying feature engineering and domain-aware matching logic to support symptom interpretation and referral routing.

- Applied GenAIOps lifecycle practices in Azure across model and output testing, evaluation, quality checks, documentation and iterative refinement, communicating assumptions and limitations for reproducible team review and handoff.

Geospatial Intelligence and Data Analysis Centre, Ministry of Defense | Abuja, Nigeria | 2020 - 2023

Pioneer Database Administrator & Geospatial Data Analyst

- Partnered with technical, operations and administrative stakeholders to define analytical requirements and translate mission challenges into geospatial, database and reporting solutions.
- Administered MySQL and MS SQL Server environments with secure role-based access controls, data-integrity checks and governed access to structured and geospatial datasets in mission-critical operations.
- Performed detailed data analysis, data wrangling, relational data modeling and anomaly investigation using Python and SQL, integrating ArcGIS, Tableau and Excel to surface location risks and decision-support insights.
- Optimized SQL queries to improve efficiency by 30%, automated recurring calculations and reporting with Excel VBA to accelerate monthly reporting by 25%, developed the centre operating manual and supervised more than 10 staff.

SELECTED PROJECTS

Project ARIEL - Deployed Generative AI Assistant | GenAIOps | Azure App Service

- Engineered an end-to-end RAG architecture with a 46-document Markdown/YAML knowledge base, 237 embedded chunks, OpenAI models, semantic retrieval and grounded response generation.
- Built version-controlled FastAPI and JavaScript components in GitHub and automated CI/CD through GitHub Actions to deploy the AI backend to Azure App Service and integrate a live GitHub Pages frontend.
- Implemented GenAIOps evaluation, monitoring and regression controls through 17 test suites, 277 audit questions, latency checks, failure analysis, patch tests and smoke tests, achieving 100% top-1 and top-k retrieval accuracy.

Commercial Building Energy Forecasting & Optimization | Predictive Modeling Lead | 1st Place

- Cleaned and merged hourly electricity, building metadata and weather data for 1,600+ commercial buildings, then engineered lag, rolling and environmental features for a future-period forecasting dataset.
- Compared XGBoost with a tuned neural network using time-based evaluation; selected XGBoost at R-squared 0.9804, MAE 5.44 and RMSE 34.28 and used SHAP and residual diagnostics to explain performance.
- Identified 10 underperforming buildings and translated 10%-20% reduction scenarios into \$858K-\$1.7M in projected annual savings, supporting audit, retrofit and peak-load recommendations in a first-place competition submission.

Visual Analytics & BI Dashboard Portfolio | Tableau | Power BI | Data Storytelling

- Built an Airline Delay Analytics dashboard by reshaping airport and carrier data into rate comparisons, interactive maps and trend views that benchmarked Buffalo Niagara International Airport against peer airports.
- Built a Stack Overflow Career Priorities dashboard by transforming workforce survey data into dumbbell charts, heatmaps and salary-segment views comparing compensation, training, diversity and office culture.
- Built a Hofbrauhaus Competitive Radius dashboard with adjustable distance filters, business-category switching and dynamic KPI counts to analyze nearby restaurants and food-store competitors from Buffalo license data.

Customer Purchase Prediction & Conversion Analytics | Product Analytics

- Built an end-to-end purchase-propensity workflow across 12,205 cleaned shopping sessions, analyzing class imbalance, visitor behavior and PageValues before leakage-safe preprocessing.
- Compared Logistic Regression, Random Forest and Gradient Boosting with consistent evaluation, selecting Gradient Boosting at ROC-AUC 0.9377 and interpreting behavioral drivers through feature importance.
- Tuned the decision threshold to 0.70, achieving 75.13% recall while using precision, F1 and confusion-matrix analysis to balance buyer capture against unnecessary campaign targeting.

RESEARCH COMMUNICATION & RECOGNITION

Certified Anomaly Detection for IoT Security | Co-Author, Research Software & Evaluation | 2025 - 2026

- Developed and evaluated Poisson-CI and negative-binomial concentration detectors across 359,296 primary 10-ms IoT windows, formalizing uncertainty-aware anomaly thresholds for real-time network traffic.
- Achieved NB-CI AUC 1.0000, zero benign false positives in the strongest configuration and approximately 448 microseconds per window, with cross-dataset validation and comparison against conformal and statistical baselines.
- Co-authored two submitted machine-learning papers and packaged reproducible Python software for feature engineering, bootstrap and DeLong analysis, runtime testing, diagnostics and reviewer-ready experiments.

Presentations: “From Probability to Protection: Poisson Inequalities for Machine Learning in Cybersecurity” (2025) | “Using AI to Address Cataloging Backlogs” (2025)

Recognition: 1st Place, Data Masters Challenge (2025) | Koessler Summer Fellowship (2025) | Dean's List Excellence Award