OLUOMACHUKWU EMENIKE

MPH Candidate (Epidemiology) | Machine Learning Engineer

Onitsha, Anambra State, Nigeria | emenikeoluomachukwu@gmail.com | +234 703 288 4809

github.com/Omahcy | linkedin.com/in/oluomachukwu-emenike

PROFESSIONAL SUMMARY

MPH candidate in epidemiology with applied machine learning experience at FlyRank AI and a background in anatomy and data science training (ALX). Builds reproducible Python data pipelines and classification models for public health and search performance data, with a focus on strict data contract validation and target leakage prevention. Experienced with the Anthropic Claude model family, prompt engineering, and Model Context Protocol (MCP).

TECHNICAL SKILLS

Programming and Tools: Python, Pandas, NumPy, scikit-learn, LightGBM, XGBoost, imbalanced-learn, Git, GitHub, Linux/Bash, VS Code, Google Colab, Google Sheets

Machine Learning: Supervised learning, Classification models, Gradient boosting, Class-imbalance handling (SMOTE), Model interpretability (feature importance), Feature engineering, Exploratory data analysis (EDA), Data contract validation

AI Tooling: Anthropic Claude model family, Prompt engineering, Model Context Protocol (MCP), Multi-turn context management

Public Health: Epidemiology, Disease surveillance modeling, Cohort analysis, Risk stratification, Demographic evaluation

Infrastructure and Operations: Netlify (hosting, DNS, HTTPS), Search intelligence, Digital operations, Anatomy and biomedical science

PROFESSIONAL EXPERIENCE

AI Fluency & Search Intelligence Intern, FlyRank AI | Aug 2026 – Present

- Built an automated Python CLI diagnostic pipeline (Pandas) that enforces data hygiene, blocks target leakage, and writes a structured markdown audit receipt for every run.
- Designed data contracts and baseline action-scoring rules, then prototyped an autonomous diagnostic agent on top of them, with zero autonomous writes by design.
- Completed 20 Anthropic Claude courses covering prompt engineering, multi-turn context workflows, tool use, and agent evaluation, and applied them while building agent prototypes.

Project Assistant & Digital Department Member, Anambra State Signage and Advertising Agency (ANSAA) | 2024 – Present

- Support project workflows and digital department operations across urban commercial corridors.
- Maintain structured tracking sheets in Google Sheets and digital records to support operational reporting and data integrity.

Corps Member, National Youth Service Corps (NYSC), Nigeria | 2022 – 2023

PROJECTS

Autonomous Organic Content Refresh Triager (FlyRank AI Agent MVP) | Python, Pandas, CLI | github.com/Omahcy/FlyRank_Internship

- **Problem:** Traffic drops are often seasonal or low-value, so the real goal was finding Page-1 pages (positions 1–10) with large impression volumes but far fewer clicks than their position should earn.
- **What I built:** A Python CLI triage agent that processed 30,000 URL performance rows and ranked content refresh candidates by missed clicks versus their SERP-position CTR benchmark.
- **How:** Enforced data contract checks that quarantine post-window leakage fields (trend_direction, impressions_last_30d); imputed missing zero-rank values with cohort medians; calculated missed-click expectations by SERP position cohort.
- **Result:** Triaged 30,000 URLs down to 2,070 high-priority candidates (top ~6.9%). The top page sat at position 2.5 with 509,252 90-day impressions but a 0.15% CTR against a 2.71% cohort benchmark, about 13,000 missed clicks recoverable with a snippet refresh. Every run emits a markdown audit receipt with SERP distortion warnings and zero autonomous writes.

Epidemiological Risk Prediction Models | Python, scikit-learn, Pandas | github.com/Omahcy/diabetes-risk-prediction

- **Problem:** Compare class-imbalance strategies and models for diabetes risk prediction, with explainability and fairness checks, to support community risk stratification.
- **What I built:** An end-to-end Python pipeline on the CDC BRFSS 2015 Diabetes Health Indicators dataset (253,680 respondents, 13.9% diabetic or pre-diabetic), with engineered features (BMI by general health interaction, chronic-condition count, weighted comorbidity risk score).
- **How:** Benchmarked Logistic Regression, Random Forest, Gradient Boosting, XGBoost and LightGBM under four imbalance strategies (raw, SMOTE, under-sampling, class weighting); used a stratified 80/20 split with scaling and resampling fit on training data only to prevent leakage; ranked features with LightGBM importance; audited sex and age-group subgroups for demographic parity and equalized odds.
- **Result:** On a held-out test set of 50,736 respondents, class-weighted LightGBM reached ROC-AUC 0.827 and caught 79% of diabetic cases (precision 0.31, F1 0.44), versus AUC 0.820 and 76% recall for class-weighted Logistic Regression, a small gain from boosting. Without rebalancing, models reached a similar AUC (about 0.83) but caught only 13–15% of cases.

EDUCATION & CREDENTIALS

Master of Public Health (MPH), Ahmadu Bello University, Zaria | Expected 2027

Data Science and Machine Learning Program, ALX | Jan 2025 – Feb 2026

Bachelor of Science (B.Sc.) in Anatomy | 2016 – 2021

CERTIFICATIONS

Anthropic Claude Course Series (20 courses), Anthropic via FlyRank AI | 2026