Yi Luo

Luoy45802@outlook.com — +1 484-542-2576 — ymo26.github.io

EDUCATION

Lafayette College Easton, PA — Sep 2022 – Jan 2026
B.A. in Mathematics & Economics (Minor: Data Science)

EXPERIENCE

Lafayette College Easton, PA — Sep 2024 – May 2025
Research Assistant (Environmental Determinants of Health, Advisor: Sayorn Chin)

- Constructed a harmonized 2000–2023 county-level dataset linking mortality outcomes to environmental exposures ($PM_{2.5}$, ozone, temperature, humidity).
- Automated data retrieval (EPA, NOAA, CDC WONDER APIs), resolved missing values, and implemented reproducible R/Python pipelines for 3,000+ counties.
- Produced annual mortality maps and regressions showing elevated cardiovascular mortality linked to $PM_{2.5}$ exposure in rural U.S. regions.

Columbia University Remote — Jan 2025 – May 2025
Research Assistant (Feature Engineering & AutoML Evaluations, Advisor: J. Yang)

- Designed end-to-end ML pipelines for messy tabular datasets using automated preprocessing, KNN imputation, and nested cross-validation.
- Benchmarked AutoML defaults versus custom feature-engineered models, achieving a mean AUROC improvement of 2–6pp across datasets.
- Produced reproducible Jupyter workflows and co-authored empirical results on feature generalizability and bias reduction.

China Galaxy Securities Co., Ltd. Beijing, China — Summer 2023
Financial Analyst Assistant

- Analyzed ETFs, mutual funds, and fixed-income products using Excel and SQL to evaluate performance and risk metrics.
- Conducted time-series analysis on asset volatility and correlation across sectors to inform portfolio allocation strategies.
- Supported senior analysts with financial modeling, report automation, and client presentation materials.

SELECTED PROJECTS

Decision Trees vs. Random Forests: Robustness on Noisy Tabular Data Jan 2025 – May 2025
Python (scikit-learn), model stability, hyperparameter tuning

- Compared decision trees and random forests under label noise, imbalance, and correlated features.
- Demonstrated 10–30% higher accuracy and stronger variance stability for RF models under simulated noise.

Urban Heat Island: Mapping & Microclimate Analysis Sep 2024 – Dec 2024
R (sf, osmdata, ggplot2, lubridate), geospatial analysis, climate diagnostics

- Integrated OpenStreetMap footprints with meteorological data to map urban temperature gradients.
- Quantified day–night temperature variations across dense zones, visualized cross-city microclimate disparities.

Election Campaign Projections Using Python Sep 2024 – Dec 2024
Python (pandas, seaborn, matplotlib), swing-state prediction, data cleaning

- Built Python pipelines to clean and visualize U.S. election data; generated heatmaps of historical competitiveness.
- Identified key swing states and margins of victory influencing the 2024 electoral landscape.

SKILLS

Programming & Data: Python, R, Stata, SQL, MySQL, LaTeX, Git/GitHub, Tableau, Excel, PowerPoint