ETHAN LEE

781-690-7469 • eyechan01@gmail.com • github.com/eyechan01

EDUCATION

Harvard University — Cambridge, MA

A.B in Computer Science, GPA: 3.934/4.00 — May 2023

A.M in Statistics, GPA: 3.669/4.00 — May 2023

Relevant Coursework: Machine Learning. Artificial Intelligence. Statistical Inference. Causal Inference. Bayesian Data Analysis. Deep Statistics. Data Visualization. Probability. Quantitative Finance. Data Structures & Algorithms.

Activities: Harvard Open Data Project (President). Harvard Computer Society (Executive Board Member). Harvard Data Analytics Group. Datamatch (Algorithms/Web Dev). Harvard Crimson.

WORK EXPERIENCE

Zest AI — Los Angeles, CA

Data Scientist — March 2024 – Present

- Building equitable XGBoost and LightGBM models (Python) and end-to-end ML pipelines for credit union loan approvals, while improving risk levels of clients' loan portfolios. Improved model performances by up to 10% (AUC) through novel implementation of hyperparameter tuning (Python) and sped up model-building processes by 3-4 times through use of Spark.

THRIVE! (thrive.industries) — Washington, D.C.

AI Engineer — February 2024 – Present

- Prototyping generative AI and synthetic data generation models in Python to improve Harvard and MIT innovation award-winning school systemic equity scoring platform. Planning for future largescale deployment to school systems.

Optiver — Chicago, IL

Trainee, Trading & Research — August 2023 – November 2023

- Completed trading training program, creating tools and ML models for forecasting volatility and underlying moves in various markets. Directly improved company pricings on derivatives via novel modeling techniques.

Capital One — New York, NY

Intern, Technology — June 2022 – August 2022

- Developed a unified testing environment in React for maintaining company software quality, using REST API calls for extracting user and software information from AWS backend.

World Data Lab — Vienna, Austria

Principal Investigator, Econometric Modeling and Policy Research — August 2021 – March 2022

- Implemented econometric modeling algorithm for identifying impactful economic and financial factors in NYC poverty data.
- Selected by Alfred P. Sloan Foundation for research grant through "extremely rigorous" review.

Republic of Paraguay, Department of Public Contracting — Asuncion, Paraguay

Intern, Machine Learning — June 2021 – September 2021

- Predicted market prices of goods/services in Paraguay's public procurement process using R and Python.
- Used linear regression, random forest, XGBoost, GAM, and elastic net regression; reduced RMSE to within 3 USD of actual prices. Final model implemented in Paraguay financial corruption investigations.

LEADERSHIP & PROJECTS

Budgeting LLM Tool — New York, NY

Initiative Leader — December 2024 – Present

- Training an LLM to create an enhanced budgeting tool making recommendations to users on how to set and adjust their spending over time, based on changes to income, portfolio, and location.

Harvard Investment Portfolio Analysis — Cambridge, MA

Project Leader — May 2022– June 2022

- Used time series analysis in Python and evaluated beta, delta, and other KPIs of Harvard investment portfolio over 20 years to understand its performance and determine its most contributing factors.

Harvard Open Data Project — Cambridge, MA

President — March 2021 – April 2022

- Led and managed 250+ members (largest data science community at Harvard) and $15k+ in funds; conducted analyses on Cambridge crime rates and COVID policy.
- With Cambridge Open Data Program, organized annual datathon; led datathon with 150+ participants on analyzing housing market and affordability in Cambridge; reports presented to Cambridge zoning and housing staff.

Philippines Job Market Analysis — Cambridge, MA

Project Member — November 2020 – December 2020

- Conducted analysis of factors of first-time salaries in Philippines, investigating effects of industry, gender, and university.
- Used linear regression, mixed effects, LASSO, and random forest models; RMSE of best model as low as 0.18.

SKILLS & INTERESTS

Technical Skills: ML/Deep Learning Models (Tree-based Models, Gradient Boosting, Neural Nets), Data Science (Pandas/Scikit-learn/Keras/Tensorflow), LLMs/NLP, AWS/S3, Python, R, SQL, React/Angular, Node/Express, C++, Java, MATLAB.

Personal Interests: Data Journalism, Sports Analysis, Personal Investing, Story Writing, Soccer, Music Composition.