

Yu-An(Aaron) Chen

412-214-2260 · [Personal Website](#) · aaronchenyuan@gmail.com · [GitHub](#) · [LinkedIn](#)

Education

Carnegie Mellon University, GPA: 3.81/4.00

B.S. in Statistics and Machine Learning · Minor in Computer Science

9/2020 - 12/2023

Pittsburgh, PA

Work Experience

Analytics Engineer | Insurify

2/2024 - Present

Insurance Aggregator (\$600M Series B Valuation)

Cambridge, MA

- Scale departmental analytics by developing new data pipelines/transformations/reports while ensuring timely and reliable delivery of all data products for strategic decision making
- Direct MLOps in customer LTV and recurring revenue estimation; led transition from earned to forward revenue reporting and successfully justified a **10x increase** in YoY monthly marketing spend
- Manage 1100+ dbt models and 120+ Airflow DAGs by conducting regular QA, developing custom alerting, and implementing refactors during hyperscale growth from **\$4M to \$42M** monthly revenue
- Standardize advertiser performance reporting to offer targeted bidding recommendations, contributed to **1.8x increase** in average bids and **11x increase** in YoY revenue
- Lead upgrade of revenue attribution framework by implementing Marketing Mix Models (MMMs), resulting in expansion of television advertising from \$100k to \$2M+ monthly spend

Teaching Assistant(s) | Carnegie Mellon University

8/2023 - 12/2023

Machine Learning Dept., Dept of Statistics & Data Science

Pittsburgh, PA

- Facilitated [Machine Learning with Large Datasets](#) (graduate/PhD-level, 80+ students) by developing assignments and exams, grading student work, hosting office hours, and directing recitations
 - Key topics: Spark, distributed ML, GPUs, hashing, deep learning optimization, dimensionality reduction
- Facilitated [Modern Regression](#) (senior/graduate-level, 230+ students) by grading student work, hosting office hours, providing feedback, and proctoring exams
 - Key topics: regression theory, model inference & diagnostics, feature & model selection, regularization

Data Science Intern | Federated Hermes

5/2023 - 8/2023

Investment Manager (\$800B+ AUM)

Pittsburgh, PA

- Engineered 120+ features from sales and CRM data (19+ million rows, ~10.5GB) with aggregation and NLP methods to determine optimal client contact strategies to increase mutual fund inflows
- Built and automated data pipelines with PySpark on Databricks to streamline ML development, achieved 93.7% accuracy in predicting purchase activity & suggesting next steps
- Segmented client base based on purchase history with regression and determined ideal combination of interactions across client types for maximizing conversion rates

Machine Learning Intern | Behavior

5/2022 - 8/2022

Addiction Recovery via AI

Pittsburgh, PA

- Built frameworks for evaluating craving-predicting ML models through F-score and K-fold cross-validation, facilitating context-specific model evaluation
- Parallelized hyperparameter tuning process of CNN and XGBoost models by incorporating multiprocessing, decreased evaluation runtime by 800%
- Employed and tested multiple feature engineering techniques (Gaussian smoothing, Autoencoding, etc.) to enhance model performance, increased cross-validation accuracy by 3.5%

Research Assistant | Learnlab

1/2021 - 5/2023

CMU Research Lab (School of Computer Science)

Pittsburgh, PA

- Developed data pipelines & reporting involving NLP, mixed-effects modeling, and hypothesis testing to reveal associations between course material presentation methods and student performance
- Led the data analysis team of Podsie, an educational start-up, and analyzed the efficacy of its products under low-sample size constraints, collectively leading to renewed lab funding in both review cycles

Skills & Personal Projects

Programming Languages: Python, R, SQL, C

Tools/Frameworks: dbt, Airflow, Databricks, Apache Spark, Redshift, AWS, sklearn, Stitch, PowerBI, Mode, Hex

Certifications: Financial Engineering, Risk Mgmt, Financial Markets, Advanced MLOps, AWS Technical Essentials

Projects: [Analytics Engineering Demo \(SQL/Python\)](#) · [Portfolio Exposure Attribution \(R\)](#) · [Optimal ETF Allocation Solver \(R\)](#) · [Neural Wavelets for Time Series Compression \(Python\)](#) · [Distributed ML with Big Data \(Pyspark\)](#)