Zimu (Tim) Zhou

Jersey City, NJ | 347-527-8722 | zz3645@nyu.edu | https://timzzm.github.io/ | https://www.linkedin.com/in/ztimz/

EDUCATION

New York University   New York, NY
M.S. *Data Science,* GPA: *3.95*   Sept. 2024- May 2026
New York University   New York, NY
B.A. *Mathematics with High Honors*, GPA: *3.78 (Cum Laude)*   Sept. 2021- May 2024
Coursework: ML *(VAEs, Unet)* | Applied Statistics *(GLM, Survival Analysis, Fourier and Wavelet)*

SKILLS & RELEVANT EXAMS

- **Skills:** Python (*np, pd, plt, XGBoost, Torch, Xarray, TensorFlow*), SQL, R, AWS, Git, MS Suite, Tableau
- **Passed Actuarial Exams:** P, FM, SRM

PROFESSIONAL EXPERIENCES

Research Intern, *InsurTech NY*   *New York* **June- August 2025**
- *Insurance Data ETL & Analytics:* Built an automated **Python ETL and analytics pipeline** to integrate multi-line loss and premium data, compute monthly **premiums** and incurred/reserve/total **losses** and ratios, and produce stratified **statistical analyses** by states, age cohorts, loss quartiles, and high-value claim counts.
- *Policy Application Approval Model:* Applied **XGBoost** to build regression tree models to predict the incurred loss based on criminal record data.

Machine Learning Engineer Intern, *AutoNavi Software Co., Ltd*   **Beijing**, June- August 2024
- *Client Branch Store Sales Prediction*: Employed **XGBoost** to build regression tree models to predict the sales of client's new branch stores with **90%** of predictions within ±20% of actual values.
- *Branch Store Site Selection*: Exploited **DBSCAN clustering** to extract key characteristics of high-volume stores and used **cosine similarity** to locate similar customer regions.
- *Faculty/Students Classification*: Utilized **XGB multi-classification tree** to differentiate among **10M+** faculties, students, and other people based on their daily routes, running on **cloud computing platform.**
- *Branch Store Road Match*: Developed a **MySQL ETL script** to match the client's branch stores with nearby segmented roads from electronic map grids.
- *Trade Area Determination*: Built a **SQL pipeline** to determine the primary trade area of **150+** branch stores based on the matched surrounding roads, employing **custom Python functions utility** for data transformation.

Data Scientist Intern, *Tencent Holdings Ltd*   **Hong Kong**, July- August 2022
- *Microloan Applicants Classification*: Used **classification tree** model to classify microloan applicants into three risk categories based on previous credit history, enhancing the **risk control** framework.
- *Actual Income Estimation*: Developed a model for **income estimation** for microloan applicants by applying separate **linear regressions** to declared and TU-estimated income, taking the minimum to mitigate overstatement risks. This approach **improved accuracy by 15%** over the previous model.
- *Automated Microloan Approval*: Applied **category-specified logistic regression** to automate the microloan approval process, reducing high-risk approvals with accuracy **73%**.

PROJECTS

Applied Math Research: Moist Convection Simulations   *New York*, Sept 2023 – May 2024
- Developed **30+ Python scripts** to simulate convection models derived from Navier-Stokes equations in 2D and 3D. Ran the simulations on **HPC**, varying resolutions, initial conditions, and boundary conditions.
- Applied **PCA and Fourier Transform** to detect key periodic features from high-dimensional time-series data generated by the simulation.

Climate Data Science Research: 3D Data Reconstructions   *New York*, May 2025- Present
- Trained **35+ VAE or Unet** on **NCAR's HPC** to reconstruct 3D climate fields from 2D predictors, experimenting with various losses: **RMSE, Wasserstein, custom weighted losses**.
- Conducted comprehensive **post-analysis** of model performance through error classification metrics (threshold-based accuracy measures), histogram-based distributional comparisons, and binned diagnostic plot.
- Built large-scale training datasets by re-gridding high-resolution satellite data to structured grids, followed by **normalization, chunking, and curation** for efficient model training.