

Xingyin Xu

201-978-7301 | xingyinxu@gmail.com | www.linkedin.com/in/xingyin-xu | <https://yinyin0916.github.io/>

EDUCATION

New York University

Master of Science in Data Science

GPA: 3.95/4.0

Relevant Courses: Machine Learning, Visualization for Machine Learning, Natural Language Processing, Computational Cognitive Modeling, Capstone, Big Data, Programming for Data Science, Practical Training for Data Science.

New York, NY

Aug. 2023-May. 2025 (Expected)

University of California, San Diego

Bachelor of Science in Probability and Statistics, Minor in Data Science and Economics

GPA: 3.98/4.0 (top 2%)

Relevant Courses: Data Analysis and Inference, Data Structures, Data Management, Machine Learning, Recommender Systems, Mathematical Statistics, Stochastic Processes, Mathematics of Finance, Econometrics, Linear Algebra.

San Diego, CA

Sept. 2019-Jun. 2023

WORK EXPERIENCE

GardenStar Group

Data Scientist Intern

Jersey City, NJ

Sep. 2024-present

- Scraped and analyzed real estate data in Hampton, GA (30228 zip code) for market trends and pricing insights.

New York Life

Data Scientist Intern

Tampa, FL

Jun. 2024-Aug. 2024

- Collaborated closely with marketing team to derive actionable **business insights** from direct mail campaigns.
- Ensuring data quality and relevance for **300+ variables** through rigorous screening and exploratory data analysis.
- Built a robust marketing targeting model for direct mail using **Logistic Regression, Random Forest, and XGBoost** iteratively, enhanced the campaign effectiveness by **100%**.

China Construction Bank

Data Analyst Intern

Suzhou, China

Jul. 2023-Aug. 2023

- Developed and implemented a **VBA-based automated system** for daily deposit reports, enhancing the detection of anomalies and inconsistencies. Reduced manual workload by approximately **50%**.
- Created and deployed an interactive data dashboard using visualization techniques (line and pie graphs, slider controls) in **Tableau**. Improved customer relationship management, leading to **40%** increase in operational efficiency.

Halcioğlu Data Science Institute – UC San Diego

Instructional Assistance for Data Structures and Algorithms

San Diego, CA

Jan. 2022-Jun. 2023

- Collaborated with faculty to craft course materials, ensured clarity and comprehensibility in assignments and exams.
- Led office hours each week and explain complex concepts into understandable segments in **Python** and **Java**.

Clear Creek Capital

Data Analyst Intern

Los Angeles, CA

Jun. 2022-Sept. 2022

- Spearheaded a data scraping project to collect **50,000+ data points** on crude oil, gold, and treasury bonds. Applied robust intermarket analysis during different macroeconomic cycles, refined profitability models and investment strategies.
- Updated existing datasets with **SQL queries** for market sentiment reports, optimizing join operations and utilizing window functions to lower runtimes and costs.

PROJECTS

Mitigating Overconfidence in LLMs

New York, NY

Finetuned Vicuna and LLaMA 2 using Verbalized Confidence Level and CoT

Oct. 2023-Dec. 2023

- Employed **Vicuna** and **prompt engineering** to generate verbalized confidence levels and answers of multiple choice and sentiment analysis datasets with **PyTorch**.
- Knowledge transferred from GPT-4 to Vicuna by finetuning, improved LLM accuracy, ratio of overconfidence bias, and ECE for **15 datasets**.
- Ensured accuracy of verbalized confidence levels and answers by data cleaning using **Python** Numpy and **OpenAi**.

Movie Ratings Analysis

New York, NY

Applied Statistical Tests and Machine Learning to Analyze Ratings and Raters

Oct. 2023-Dec. 2023

- Utilized **Mann-Whitney U tests** and found **significant** rating difference on engaged and not engaged audience.
- Built **Linear Regression Model** to predict movie ratings and utilized **grid search** to build **Lasso regularization**.
- Examined quality consistency of franchise movies using **Kruskal-Wallis H-test**, and utilized consistent franchise movies ratings to build regression model. Improved R square by more than **150%**.