

XIAOXUE XIONG

xx2385@columbia.edu | 929-665-7649 | New York, NY | [Github](#)

EDUCATION

Columbia University **New York, US**

M.A. in Statistics GPA: 3.97/4.00 *Sep 2022 - Feb 2024*

Coursework: Advanced Machine Learning, Time Series, Natural Language Processing, Databases, A/B Testing, etc.

Prizes: Kaggle Tensorflow - Help Protect the Great Barrier Reef (top 2% out of 2025 teams, 2022)

Zhejiang Gongshang University **Zhejiang, China**

B.S. in Statistics GPA: 89/100 *Sep 2017 - Jun 2021*

Coursework: Probability, Nonparametric Statistics, Machine Learning, Statistical Inference, Data Structure, etc.

Prizes: Kaggle Google Brain - Ventilator Pressure Prediction (top 7% out of 2100 teams, 2021) & Meritorious Winner in the Mathematical Contest in Modeling (Top 8%, 2020)

EXPERIENCE

Indicatorlab **New York, US**

Quantitative Trading Analyst *Feb 2024 - Present*

- Collected data from API or by web crawlers, built ETL pipeline, updated to NoSQL database and optimized strategies
- Developed interactive dashboards in Tableau to visualize key performance metrics, enabling data-driven decision-making
- Built the model training and inference modules in Python, emphasizing modularity, reusability, and efficiency
- Delivered results in a fast paced environment using Microsoft Office Suite (Excel, Word, PowerPoint)

JD.COM (*Top-tier E-commerce in China*) **China**

Data Scientist Intern *Sep 2023 - Nov 2023*

- Utilized SQL to extract insights from structured and unstructured data, presented findings to stakeholders
- Enhanced personalized recommendations by implementing collaborative filtering, k-nearest neighbors and SVD
- Diagnosed business needs and developed ML models tailored to optimize users' best actions towards items (add to cart/buy/bid/watch/save), thereby significantly enhancing user engagement

Samara Alpha Management **New York, US**

Quantitative Trading Intern *May 2023 - Aug 2023*

- Extracted insights from millions of trading records by exploring different feature engineering, data pre-processing techniques, and evaluating different ML models like Lasso, LightGBM, etc
- Designed and implemented systems to monitor and set up alerts for real-time news
- Presented findings to stakeholders using Microsoft PowerPoint

Xizhi Credit Risk Management **China**

Data Scientist *Sep 2021 - Jun 2022*

- Built extract, transform, load (ETL) pipeline to read/load data from internal data sources and wrote automated analysis code to provide reusable analysis framework (SQL/Python)
- Created a NLP (text classification) model by analyzing client feedback data to boost operational efficiency and reduce expected annual cost of complaints from client churn by 10%
- Designed a new metric, financial health score, by analyzing historical client data to help optimize individuals' scores
- Developed PD (probability of default) models to boost model performance (KS score) by 20%, by using feature engineering and XGBoost model

Alibaba Group (*Top-tier E-commerce in China*) **China**

Data Scientist Intern *Jan 2021 - Mar 2021*

- Analyzed A/B testing results in hypothesis testing, extracted causality between UI change and customer conversion rate
- Utilized Python, R, and SQL to conduct data wrangling, exploratory data analysis, and statistical modeling
- Presented results to team leaders using Microsoft Office Suite

RESEARCH

Columbia University **New York**

■ **Cointegration-based Production Spread Trading** *Sep 2023 - Nov 2023*

- Designed a trading strategy to arbitrage on production spread (WTI Crude Oil vs. RBOB Gasoline)
- Applied Cointegration Test to weight oil and gasoline in this strategy, achieving an Information Ratio over 0.8

■ **TimeInf: Time Series Data Contribution via Influence Functions** *Sep 2023 - Feb 2024*

- Proposed TimeInf method, utilizing influence functions to attribute model predictions to individual time points while preserving temporal structures
- TimeInf outperforms the state-of-the-art methods in identifying harmful anomalies and helpful time points for forecasting

SKILLS

- **Programming & Tools:** Python, SQL, R, MATLAB, A/B Testing, Microsoft Office Suite (Excel, Word, PowerPoint)
- **Data Visualization:** Tableau, Python, R, Power BI
- **Database & Big Data:** SQL (Microsoft SQL Server, PostgreSQL), NoSQL (MongoDB), GCP, AWS (S3)