

Jingyi (Alina) Zhou

619-937-1899 | alinazhoucareer@gmail.com | [linkedin.com/in/alinazhou-tjuucsd](https://www.linkedin.com/in/alinazhou-tjuucsd)

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY	Berkeley, CA
<i>Master of Analytics</i>	GPA: 4.0/4.0 (Freshman Scholarship) Expected August/December 2025
UNIVERSITY OF CALIFORNIA, SAN DIEGO	San Diego, CA
<i>Master of Science in Data Science</i>	GPA: 4.0/4.0 (Academic Excellence Award) June 2024
TONGJI UNIVERSITY	Shanghai, China
<i>Bachelor of Engineering with a Minor in Accounting</i>	GPA: 3.7/4.0 (Outstanding Scholarship) July 2022

SKILLS

- *Technical Skills:* Python, MySQL, R, Spss, Neo4j, MongoDB, PostgreSQL, DataWorks, Matlab, Spark, Tableau, Linux
- *Courses:* Recommender Systems, Text Mining, Causal Inference, Machine Learning, Optimization, and more

PROFESSIONAL EXPERIENCE

ALIBABA GROUP	Shanghai, China
<i>Data Science Intern, User Growth Team of Ele.me E-commerce Platform</i>	June 2024 – September 2024

- Developed and integrated user labels from multiple data sources by utilizing **SQL** to process 900 million records and optimized the efficiency through **resource allocation** and **query tuning**, improving processing speed by 30%
- Addressed user label distortion issues by innovating new features and employing an **XGBoost** model to achieve 89.3% accuracy and 86% MAC recall, conducted **offline evaluation** using key metrics, and built a **data dashboard**
- Innovated user consumption segment using new labeling methods, supported by **A/B testing**, lifting order frequency by 3%, which directly informing operational strategies launched in September 2024
- Collaborated with product, operation, and BI teams in developing user labels, achieving a daily usage of over 1 billion API calls and enabling seamless integration with more than 600 downstream tables
- Analyzed the growth potential of users by utilizing **data visualization** and conducting detailed **metric drill-downs**

CHINA CHENGXIN INDICES	Beijing, China
<i>Financial Data Science Intern, Financial Engineering Group</i>	January 2022 – December 2022

- Developed **ETL pipelines** with **Python** and **SQL** to process 1 billion financial records from multiple data providers and implemented multiple partitioned tables in DWS and ADS layers to extract features and support downstream analytics
- Utilized **PyTorch** to construct **CNN** models on stock candlestick charts to predict stock trends, achieving a precision of 61% and improving the effectiveness of an established trading strategy by 5%
- Innovated a **CNN** model incorporating six time-series feature operators for **automated feature engineering** to predict stock returns, resulting in 22.3% annualized return, which was presented at the *CFRI & CIRF Joint Conference 2023*
- Developed a sentiment classification model by **fine-tuning** the **FinBERT** model on manually labeled corporate earning calls, achieving an average F1 score of 0.88 in a three-class setup

RESEARCH & PROJECT

Research on Causal Inference for Online Advertising	March 2024 – June 2024
--	------------------------

- Developed an **uplift modeling** framework using **meta-learning** and **direct modeling** methods to measure advertising lift effects and identify customer segments most responsive to advertising interventions
- Explored various meta-learning methods and algorithms, with **T-learner** approach achieving the best AUUC of 0.83
- Built causal graph models using **Graph Neural Networks**, enhancing accuracy with node embeddings, and evaluated robustness by testing causal graph generation methods with random edge-flipping strategy, achieving 0.87 AUUC

Credit Card Fraud Detection Model	March 2024 – June 2024
--	------------------------

- Conducted **Data Quality Review**, **data cleaning**, and **EDA** on 97,852 credit card transaction records, developed 3,337 transaction behavior features, and identified the top 20 features using **filter and wrapper** methods for **feature selection**
- Achieved a 90% fraud detection rate for top 20% risky groups with a 10% false positive rate with **LightGBM** model
- Selected a threshold by testing financial saving curves, achieving 61.95% FDR and estimated savings of \$44.3 million

Statistical Analysis of River Water Quality Monitoring Data	September 2021 – January 2022
--	-------------------------------

- Analyzed water quality trends through **correlation analysis**, **regression analysis**, and **cluster analysis** in **Scikit-learn**
- Identified seasonal water quality patterns, concluding that the river qualifies as Class-I surface water and that water quality from June to September is most suitable for fish through **t-test**, **cross-tabulation**, and **Chi-square test**
- Developed a data-driven framework that integrates water quality indicators into a scoring system, enabling local authorities to rank high-priority areas for water conservation efforts and allocate resources effectively