

Yishan Cai

yic075@ucsd.edu | (858) 247-9139 | [LinkedIn](#) | La Jolla, CA

EDUCATION

University of California, San Diego

La Jolla, CA

B.S. Data Science and Probability & Statistics

Sep. 2021 – Jun. 2025 (expected)

- **GPA:** 3.81/4.0, Provost Honors
 - **Relevant Coursework:** Data Structure and Algorithms, Machine Learning, Data Mining, Database Management, Data Visualization, Scalable Analytics Systems, Math Statistics, Stochastic Process, Multivariable Calculus, Linear Algebra
-

SKILLS

- **Languages:** Python, SQL, Java, R, MATLAB, JavaScript, HTML, CSS
 - **Libraries & Tools:** NumPy, Pandas, Scikit-learn, Matplotlib; Tableau, Power BI, Stata, ArcGIS
-

WORK EXPERIENCE

Fourth Paradigm Technology Co., Ltd.

Beijing, China

LLM Data Intern, Science and Technology Department

Jun. 2023 – Sep. 2023

- Performed data cleaning on large-scale unstructured data, including Zhihu articles, Toutiao articles, etc., by using python, and obtained 923G data
 - Evaluated various indicators and characteristics of cleaned corpus data, completed over 45 evaluation reports of tens of thousands of data, and improved average accuracy by ~22% after optimization
 - Developed 2000+ prompts, and originated corresponding answers for supervised fine-tuning (SFT) model training
-

RESEARCH & PROJECTS

Chinese Academy of Agricultural Sciences, Institute of Vegetables and Flowers

Aug. 2023 – Sep. 2023

Research Assistant

Advisor: Shuhua Yang

- Implemented Principal Component Analysis (PCA), and developed Score Plot and Loading Plot to analyze the contribution value of different indicators of Feng-Hua Rose
- Derived correlation heatmaps using R to visualize the effects of three different arbuscular mycorrhiza fungi (AMF) on the morphology and physiology of rose

Bank Account Fraud Detection | [report](#)

- Analyzed the Bank Account Fraud (BAF) dataset, conducted EDA, statistical testing, feature selection, and realized Matplotlib visualizations
- Employed and optimized predictive models using Decision Tree, XGBoost, and CatBoost algorithms to detect fraudulent bank account applications, achieving 80% accuracy
- Conducted in-depth feature engineering and hyperparameter tuning using GridSearch with 5-fold cross-validation, leading to the identification of crucial features that contributed to enhanced model performance

Rainfall and Mental Health Data Analysis

- Utilized API requests for data acquisition, conducted data cleaning and wrangling, and applied inferential analysis techniques such as OLS regression and hypothesis testing to evaluate variable relationships' statistical significance.
- Applied support vector regression (SVR) for depression prediction using precipitation levels, and evaluated via MSE

Interactive 10-Year Treasury Yield Visualization

- Developed the structure, style, and design of the [website](#) using Svelte, HTML, and JavaScript
 - Examined and interpreted 10-year treasury yield dynamics and economic indicators, providing actionable insights and clear explanations for informed financial decision-making
-

LEADERSHIP EXPERIENCE & VOLUNTEERING

Chinese Students and Scholars Association, UCSD

Mar. 2022 – Mar. 2024

Head of HR Department

- Organized Spring Festival Gala for two consecutive years, attracting an audience of more than 1400 attendees. Invited 5 rappers, 7 social media influencers, 6 collaborating associations, and over 10 professors to participate
- Devised 3 organizational articles regarding experiences of new freshman on campus, reaching over 1000 readers on average
- Recorded attendance of all departments, and negotiated any issues between officers

International Students & Program Office, UCSD

Jun. 2022 – Sep. 2022

Mentor

- Guided seven international mentees on the various facets of college life from adjustment to academics to campus life
- Wrote weekly emails, and held Q&A at zoom to provide friendly support for more than 5 hours per week