

Shuaiheng Tao

820-758-2310 | shuaihengtao@gmail.com

EDUCATION

University of California, Los Angeles

Bachelor of Science (B.S.)

- Double Major: Statistics and Data Science & Mathematics
- Cumulative GPA: 3.79/4.0

Los Angeles, CA

September 2023 - March 2025

University of California, Santa Barbara

- Completed coursework in Statistics and Data Science & Mathematics

Santa Barbara, CA

September 2021 - June 2023

PROFESSIONAL EXPERIENCE

JD.COM

Data Scientist Intern, JD.com Technology Division

Beijing, China

July - September 2024

- Extracted features from thousands of URLs in emails and deployed an XGBoost model in Python to classify them as legitimate or phishing, achieving a 97% accuracy and F1 score, and reducing alert noise.
- Cleaned and preprocessed text data to train a BERT model for detecting malicious responses from large language models, achieving 94% accuracy and outperforming ChatGPT-4o with an 81% accuracy.
- Implemented the BERT model as an additional validation layer with ChatGPT-4o, ensuring that only responses approved by both models were classified as “not malicious” to maximize the recall.

Hengzheng Tax Agency

Data Analyst Intern

Huizhou, China

June - August 2023

- Developed a Python-based data-processing pipeline utilizing Pandas and regular expressions, automating the transformation and integration of data from diverse Excel sheets, significantly improving workflow efficiency.
- Conducted exploratory data analysis (EDA) and created impactful data visualizations to provide actionable insights into clients' tax reporting, supporting compliance and strategic planning.

PROJECTS

California Fiscal Health Analysis

October 2024 - November 2024

- Analyzed fiscal risks and challenges faced by California cities to promote economic efficiency, mitigate potential mismanagement, and prevent waste, fraud, or abuse in Python.
- Conducted EDA to identify key financial trends and risk factors, presenting actionable insights to stakeholders.
- Developed and implemented a random forest model to predict fiscal vulnerabilities, enhancing decision-making processes and resource allocation.

Regional Unemployment Rate Analysis

May 2023 - June 2023

- Conducted an in-depth analysis of the relationship between regional unemployment rates and various factors.
- Implemented Principal Component Analysis (PCA) for dimensionality reduction and performed simple and multiple linear regression with statsmodels and scikit-learn.
- Created visualizations of the results, including the PCA plot, kernel density estimations, and correlation heatmaps, using Altair and Matplotlib.

ADDITIONAL INFORMATION

Programming: Python (NumPy, Pandas, statsmodels, scikit-learn, PyTorch, Matplotlib, Altair), R (tidyverse, dplyr, caret, ggplot2), SQL, C++

Extracurriculars:

- President, UCSB Go Club: Won the 2023 California State Go Championship.
- Member, UCSB Data Science Club and UCLA Statistics Club.