Zelong/Alan Wang

Contact

Address

La Jolla, CA, 92122

Phone

858-539-5564

E-mail

zew013@ucsd.edu

GitHub/Personal Web

zew013.github.io

Test Taken

GRE Verbal: 162

GRE Quantitative: 169

Scholarship & Honor

TRELS $1000 Scholarship

Two-year Provost Honors

Skills

Python — Proficient

Stata — Very Good

R — Good

Microsoft Office Suit — Good

SQL — Still learning

Java — Average

HTML — Average

Git/Github — Average

Tableau — Average

MATLAB — Some Experience

Mandarin — Excellent

Time management — Excellent

Python Packages

Pandas, NumPy, bs3
Matplotlib, SciPy, Plotly,
Folium, Scikit-Learn,
PyTorch, etc.

Education

The University of California, San Diego | 2020-Present | GPA:3.98

B.S | Data Science

Related Coursework: Python, Java, HTML, Data Structures, Data Science Algorithms, Web Scraping, Data Mining, Visualization, Machine Learning, NLP basics, Deep Learning, Recommender System, Network Science & Graph Theory.

Note: Will learn big data and systems such as MapReduce/Hadoop and Spark systematically this Winter. I would have some good understanding and practice before next summer.

B.S | Joint Major in Math and Economics

Related Coursework: Calculus, Linear Algebra, Numerical Analysis, Optimization, Probability, Statistics, Micro & Macroeconomics, Econometrics (t, z, Chi-squared, F test, Regressions, panel data, etc.), Operations Research.

Experience (Sorted by starting time)

Researcher | Professor Dale Squires and Professor Richard Carson | July 2022 – Present

- Explore patterns in terms of international organizations and their membership fees to determine Social Discount Rates.
- Automatically extract and prepare data from low-quality UN reports (1950-2022) using Python, some APIs, and AWS.

Tech Consulting Intern | IBM Consulting, China | June. 2022 – August. 2022

- Provided reports about a consolidated cement company for the team. Collaborated with cross-functional stakeholders, developed scalable solution and worked towards consensus across the organization. Built metrics to inform performance.

CSE-PACE Program Designer | UCSD CSE Department | Apr. 2022 – Sep. 2022

- Searched and reviewed a miscellaneous assortment of papers, articles, programs, and games. (approx. 30 pages/week)
- Design CSE cohort programs and build 7 courses on various CS/DS related topics.

Research Assistant | UCSD Professor Richard Carson | Dec. 2021-Present

- Did literature review and gained background knowledge on discrete choice models and related topics.
- Arranged and corrected research data to create representative graphs intended for academic paper. (approx. 60 drafts)
- Coded Monte Carlo Simulation, conditional logit (max/minimum Gumbel assumption). Performed qualitative analysis.

Data Analyst/Tech VP | Lumnus Consulting (Student Enterprise) | Nov. 2021-Present

- Collaborated with other leaders to identify and prioritize problems. Organized team events and fostered team communication. Responsible for website maintenance using React.js and Heroku.
- Built prediction models based on historical Instagram and Twitter data. Created visualizations and analyses for the team. Delivered project presentations on how we should present our organization on social media.

Tutor & Math Team Leader | 2019 - 2021

Data Manager & Builder | Robotics Team | 2018 – 2020

Self-Driven Projects

Logistic Regression and Softmax Regression | NumPy, Git

- Use only **NumPy** package to implement **Mini-Batch Gradient Descent**, Cross Validation, achieved Softmax Model.

Forage Data Analytics Virtual Experience programs | Python, Tableau, Excel, Git

- "Accenture": Data mining in **Python**, created **Tableau** dashboard. Storytelling and **presentation**.
- "BCG": Deployed automatic EDA using Python **SweetViz**. Defined, assigned price sensitivity to each customer. Tested hypothesis by **permutation test**, **K-S test**, and heat map. Predicted customer churn by **Random Forest** and **XGBoost**.

Analysis of Power Outage Status in the Continental U.S. | Python, Excel

- Went through the full process of questioning, researching, data cleaning & EDA, missingness assessment, hypothesis test, baseline model, **scikit-learn** ML pipelines, final model, and fairness analysis.

Indian Crop Production and Indian Climate Analysis | Python, Slides

- Read reports about Indian agriculture, pastoralism, and climates. Reported findings and provided suggestions.
- Data gathering, data mining, interactive visualization using **bs4**, **JSON**, **Geopandas**, **Altair**, **Folium**, **Plotly**.

Simple Language Model | Python

- **Web scraping** for specific books from a public book website. Tokenize corpora.
- Created an **N-Gram Language Model** that can generate paragraphs resembling the style of an author or a book.

Wealth Prediction based on 1991 Survey of Income and Program Participation (SIPP) | R

- Applied EDA, feature engineering. Incorporated and compared **polynomials** and **splines** in **GAM**. Tested **full OLS**, **Lasso**, **Ridge**, **Stepwise regression model** and **Random Forest model**.