Weijie Zhang

wzhang3912@gmail.com | 747-283-5852 | linkedin.com/in/wzhang3912 | Green Card Holder

Education

University of California, San Diego, Sept 2022 – June 2026
Bachelor of Science in Data Science, Minor in Mathematics

- GPA: 4.0/4.0
- **Coursework:** Probabilistic Modeling & Machine Learning, Deep Learning, Database Management, Data Visualization, Data Structure & Algorithm, Probability & Statistics, Optimization, Linear Algebra

Technical Skills

Programming: Python (Scikit-learn, PyTorch, TensorFlow, Pandas, NumPy, Matplotlib, Seaborn, Plotly, BeautifulSoup), SQL, Java, JavaScript, D3.js, HTML/CSS
Data Analysis: Data Wrangling, Data Visualization, Hypothesis Testing, Machine Learning
Tools & Technologies: Git, PostgreSQL, Tableau, Jupyter Notebook, Excel

Projects

Recipe Rating Prediction with Imbalanced Data | *Python (Pandas, Plotly, Scikit-learn)*

- Performed modularized **data cleaning** processes on numerical and textual data, ensuring data integrity for analysis
- Conducted **exploratory data analysis** (EDA) and **hypothesis testing** to extract data insight and address the core research question
- Built a comprehensive **modeling pipeline**, from data preprocessing, feature engineering, and model selection, to hyper-parameter tuning, increasing overall **f1-score by 20%** compared to the baseline model

Google Local Review Recommender System for Business | *Python (TensorFlow, Pandas)*

- Implemented data pipeline for **data loading, preprocessing, and feature engineering** in batch processing for 400k+ records
- Built and fine-tuned **Temporal Dynamic Latent Factor Model** and **Factorized Personalized Markov Chain recommender model**, achieved enhanced recommendation accuracy, by leveraging TensorFlow for scalable performance

LLM-Driven SQL Database Tool | *Python, Postgres, SQL, Large Language Model, Prompt Engineering*

- Developed LLM-powered SQL query interface for "chatting" with PostgreSQL databases by automating SQL generation and natural language answering, significantly reducing query formulation time
- Engineered pipeline that auto-generates database schemas for new connections and flags potentially destructive SQL commands
- Utilized **prompt engineering techniques** to optimize model output and integrated both open-source and closed-source models with **real-time streaming**

Interactive Data Visualization Web Development | *JavaScript, D3.js, HTML/CSS*

- Developed an interactive data visualization website explaining the Laplacian Eigenmaps Algorithm using **HTML, CSS, JavaScript, and D3.js** tech stack, enabling users to explore the algorithm through dynamic, visually engaging visualization features

Experience

Instructional Assistant Sept 2024 – Present
Halıcıoğlu Data Science Institute - UC San Diego, La Jolla, CA

- Host weekly office hours, providing individual and group support to students for understanding **data science course concepts** and clarifying homework assignments
- Actively engage with students on online discussion forums such as Campuswire and Piazza, providing timely and detailed answers to student's questions