

Weijie Zhang

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

Education

University of California, San Diego,
Bachelor of Science in Data Science, Minor in Mathematics

Sept 2022 – June 2026

- 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 **data cleaning** processes on both numerical and textual data with **exploratory data analysis** and **hypothesis testing** to uncover key drivers of recipe popularity, providing insights for targeted marketing
- Developed a **predictive model** that forecasts recipe ratings by building a comprehensive modeling pipeline, from **data preprocessing**, and **feature engineering**, to **hyperparameter tuning**, increasing the **F1-score** by **20%** compared to the baseline model
- Addressed **data imbalance** by integrating oversampling and undersampling techniques on the RandomForest model, ensuring robust performance and reducing bias in prediction outcomes

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

- Implemented data pipeline for **data loading, preprocessing, and feature engineering** of over 400K Google reviews records in batch processing
- Built and fine-tuned recommender systems with **collaborative filters** and **content-based filtering** models using TensorFlow, significantly improves recommendation accuracy to drive downstream business application

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

- Developed LLM-powered SQL query interface for "chatting" with the databases by automating SQL generation and question answering, significantly reducing query formulation time when accessing the database
- Utilized **prompt engineering techniques** and integrated both open- and closed-source models with **real-time streaming** to ensure smooth and robust user experience
- Engineered automated pipeline for dynamic database schemas generation and safety checks, migrating risks from destructive queries and enhancing system reliability

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

- Designed and developed an interactive 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

Halicioğlu 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