

Stephanie Wang

(510)-918-8739 | ssyuw0418@gmail.com | ssyuw.github.io | U.S. Green Card Holder

EDUCATION

University of California, San Diego

Sept. 2022 – (expect) Mar. 2026

Bachelor of Science in Data Science | Minor in Mathematics and Cognitive Science

GPA: 3.8/4.0

- Coursework: Probabilistic Modeling & Machine Learning, Representation Learning, Data Structure & Algorithm, Search Algorithms, Prob & Stat, Data Visualization, Mathematical Reasoning, Linear Algebra, Vector Calculus

TECHNICAL SKILLS

Languages: Python (pandas, numpy, scikit-learn, xgboost, lightgbm, catboost, seaborn, matplotlib), Java, SQL (Postgres), JavaScript, HTML/CSS, R, JSON

Technical: Machine Learning Models (Tree Models, Linear Models, Naive Bayes, Ensemble Methods), Deep Learning (PyTorch, TensorFlow), Data Analysis (Exploratory Data Analysis, Feature Engineering, Data Cleaning), Statistical Analysis (A/B Testing, Hypothesis Testing), Data Visualization (jQuery, D3, Tableau), Web Scraping

EXPERIENCE

Chuling Information Technology Co., Ltd.

Jul. 2024 – Sept. 2024

Data Scientist Intern | *Machine Learning, Feature Engineering, Time Series Data*

- Cleaned, integrated, and analyzed internal sales data to improve data consistency and quality, ensuring accurate insights for business decisions.
- Developed predictive models using **Gradient Boosted Decision Trees (GBDT)** and **time series analysis** to forecast customer behavior and intent, optimizing sales strategies.
- Generated detailed data reports and collaborated with the team to implement data-driven decisions, resulting in improved customer conversion rates.
- Leveraged Python to process data and build interactive dashboards to effectively visualize analysis results for stakeholders.

PROJECTS

Rating Prediction for Amazon Products | *Machine Learning, Natural Language Processing*

- Processed and cleaned 490k+ records, applying imputation, text normalization, and outlier detection to improve data quality.
- Engineered **NLP** features with **TF-IDF**, **Word2Vec**, and **Doc2Vec** to enhance sentiment and rating prediction.
- Built and optimized a **logistic regression model**, optimizing performance with L1/L2 regularization and feature selection.
- Assessed model accuracy using RMSE and R-squared, performed error analysis, visualized key insights and feature importance to support data-driven product recommendations.

Power Outage Risk Analysis Modeling | *Machine Learning, EDA, Feature Engineering*

- Conducted exploratory data analysis to uncover patterns and trends in geographic, climate-related, and economic data related to power outages.
- Designed and implemented feature engineering processes, including encoding time series data and extracting key insights from geographic and temporal variables to enhance model performance.
- Built and optimized a predictive model using **gradient boosting trees**, achieving high accuracy in identifying power outage causes linked to natural disasters.
- Performed hyperparameter tuning by analyzing recall, precision, and **F1 scores**, improving the model's overall reliability and predictive capability.

LLM Based Web Agent | *Large Language Model, Prompting Engineering, Computer Vision*

- Automated browser-based tasks by implementing dynamic page interaction and data extraction workflows using Selenium.
- Applied **OpenCV** to perform background subtraction and boundary detection, enhancing the accuracy of target detection and feature extraction processes.
- Integrated **large language models (LLMs)** into the system, designing and fine-tuning prompts to automate complex tasks while optimizing inference speed and accuracy.