Derek Wen

San Diego, California · 858-437-3576 · US Citizen
derek.hz.wen@gmail.com · LinkedIn · Website

PROFESSIONAL EXPERIENCE

Claritas, San Diego, CA — April 2023 – February 2024
Data Science Intern

- Developed a scalable PySpark pipeline, filtering over 1 million IP addresses across diverse data sources, increasing data processing efficiency by 30%.
- Automated the extraction of campaign data from Parquet files and calculated daily data quality metrics, reducing manual effort by 20%.
- Conducted multidimensional analysis of campaign data across tables and dates, identifying trends and detecting anomalies to improve data-driven decision-making.

Nurlink, San Diego, CA — July 2021 – Aug 2021
Data Analyst Intern

- Created an automation tool to clean and preprocess circuitry data, enhancing visualization accuracy and efficiency in JMP software for circuit bench datasets.
- Ran analysis on features such as inductance and capacitance to measure their correlation with circuit KPI's like output power. Dimensionality reduction techniques such as PCA were used to conduct 2-D visualization of noise in output power.

EDUCATION

University of California, Los Angeles — 2026
Masters of Applied Statistics and Data Science

University of California, San Diego — 2024
Bachelor of Data Science + Minor in Cognitive Science
GPA 3.98 / 4.0

- Awards: Summa Cum Laude

Canyon Crest Academy — 2020
GPA 4.45 / 4.0

- Awards: Provost Honors, President's Volunteer Service Award, AP Scholar with Distinction

SKILLS & PROJECTS

Skills: Python, SQL, Apache Spark, PySpark, PostgreSQL, SQLAlchemy, PyTorch, Pandas, NumPy, R, Java, Javascript, ArcGIS, D3, MatLab, R, SPSS, Flask, Chart.js, and JMP.

Projects

- Developed and fine-tuned a DETR-based computer vision model to detect and classify utility poles from over 800 Google Street View images, achieving high accuracy in differentiating material types and automating utility pole monitoring for wildfire prevention.

- Created a RandomForestClassifier model to predict stock trades (buy/sell) by U.S. House members using trade data and representative attributes, achieving a 65% test accuracy through hyperparameter tuning with GridSearch. Conducted a fairness analysis, revealing bias in model predictions between political parties.

- Developed Matchify, a Flask-based web application integrating Spotify's API to visualize and compare users' top artists and tracks. Implemented secure authentication, interactive data visualizations with Chart.js, and deployed on Heroku with custom domain support.