DEREK WEN

San Diego, California · 858-437-3576

derek.hz.wen@gmail.com · LinkedIn · Website

PROFESSIONAL EXPERIENCE

Claritas, San Diego, CA   April 2023 – Present
Data Science Intern
- Designed and implemented a scalable PySpark pipeline to filter and process over a million IP addresses from diverse data sources within a distributed data environment.
- Developed a Spark-based pipeline that extracts campaign data from Parquet files, processes it, and calculates the daily percentage of data quality warnings across multiple metrics and tables.
- Analyzed the data across various dimensions, including tables and dates, to understand trends and anomalies in campaign data.

Nurlink, San Diego, CA   Juy 2021 – Aug 2021
Data Science Intern
- Worked on an automation tool that cleans and preprocesses circuitry data in preparation for loading into JMP software for data visualization.
- Built data visualizations using JMP software on circuit bench datasets. Ran analysis on features such as inductance and capacitance to measure their correlation with circuit KPI's like output power. Also, dimensionality reduction techniques like PCA were used to conduct 2-D visualization of noise in output power.

EDUCATION

University of California San Diego   2024
Bachelor of Data Science
GPA 3.98 / 4.0
- Member of the Data Science Student Society
- Relevant Coursework: Data Structures and Algorithms, Theoretical Foundations of Data Science, Data Management and Scalable Systems, Visualization and Statistical Methods, Machine Learning and Probabilistic Modeling, Recommender Systems and Web Mining, Data Analysis and Modeling

Canyon Crest Academy   2020
GPA 4.45 / 4.0
- Awards: Provost Honors, President's Volunteer Service Award, AP Scholar with Distinction
- Volunteer at San Diego County's Library

SKILLS & PROJECTS

Skills: Python, Pandas & NumPy Library, Apache Spark, PySpark, Java, Javascript, D3, MatLab, SQL, PostgreSQL, R, SPSS, and JMP

Projects
- Trained and fine-tuned a computer vision model (DETR) to detect power lines and underhead structures from Google Street View images gathered using Google's API and python.

- Developed a RandomForestClassifier model to predict stock trade decisions (buy or sell) of U.S. House of Representatives members based on trade characteristics and representative attributes. The model was refined through GridSearch for optimal hyperparameters and included a fairness analysis to assess model performance across political party affiliations.

- Developed predictive models for Amazon product ratings using textual analysis and temporal factors from a 500,000-entry dataset, providing insights into review patterns and influences.