Jason Gu

jaesongu@gmail.com • (626) 277-8952 • LinkedIn • Website • Github

EDUCATION

University of California, San Diego | GPA: 3.6 | B.S. Data Science — Expected Graduation: March 2025

Relevant Coursework: Practice of Data Science, Statistical Methods, Data Science Theory, Machine Learning, Optimization, Natural Language Processing, Data Management, Systems for Scalable Analytics, Data Visualization, Reinforcement Learning, Linear Algebra, Differential Equations

WORK EXPERIENCE

Data Science Intern | Mercury Alert AI - *San Diego, CA* — Jun 2023 - Oct 2023

- Automated the data manager system with **AWS** (**DynamoDB**, **S3**, **Lambda**) to effectively self-store incorrect predictions made by the company's physical state prediction model, enhancing system reliability by 18%.
- Engineered the internal **device monitoring dashboard**, enabling real-time reporting of temperature, empty frames, and device capture errors.
- Utilized the created dashboard to ensure the stable performance of 25 devices before the shipment to the **largest order** in company history.
- Designed the data manager system to self-push relabeled data from **Jupyter Notebook** to the AWS lambda training queue, **increasing training efficiency by 50%** and optimizing model performance.

Data Analyst Intern | Redrock Biometrics - *San Francisco, CA* — Jun 2022 - Sep 2022

- Executed data analysis on 1,200 different palm models using **Python**, **Pandas**, **Matplotlib**, **and NumPy**, resulting in solutions that significantly reduced the False Rejection Rate of the PalmID system by 63.6%.
- Authored a comprehensive report detailing the methods to improve the original PalmID system and directly presented the findings to the CTO of Redrock Biometrics, which led to the implementation of my reported findings.

PROJECTS AND LEADERSHIP EXPERIENCE

Recipes Protein Analysis and Predictor | Data Science / Machine Learning Project (https://jingchenggu.github.io/Protein_Prediction/)

- Conducted EDA with 83k+ recipes data and 730k+ reviews data within **Pandas** framework regarding the correlation between protein content and average ratings and cooking times of recipes to understand the factors influencing different protein levels.
- Improved the **R-squared value of the baseline model by 250%** by transitioning from a Linear Regression model to a **Random Forest Regressor**, lowering the RMSE from 24.7 to 8.4 by incorporating a broader range of nutritional features and employing **GridSearchCV**.
- Delivered a comprehensive presentation to a UCSD health & fitness club, assisting members in identifying optimally balanced protein diets tailored for bodybuilding and weight lifting.

Academic Performance Analysis at UCSD | Data Science Research Paper (github link)

- Performed a data-driven analysis of student grades at UCSD to identify factors influencing academic performance, utilizing multiple regression models in **Scikit-Learn** to quantify the impact of predictors such as class size and departments, achieving an R-squared value of 0.702.
- Enhanced prediction model performance through **feature engineering**, creating 5 new variables and converting 3 categorical variables for effective inclusion in regression models, resulting in improved interpretability and accuracy that enabled UCSD to optimize class planning, improve resource allocation, and ultimately optimized student performance at the data science department by 30%.

Worldwide Covid Investigation | Data Visualizations Using Svelte and D3.js (https://jingchenggu.github.io/Covid_Worldwide_Website/)

- Created a visualization platform using **Svelte** and **D3.js library** for users to interact with a global map with a timeframe slider to showcase the rapid spread of COVID-19 across the world during peak quarantine time.
- Implemented a dynamic line graph for the counts of total cases, recovered, and deaths based on the position of the timeframe slider.

Evil Geniuses Social Media Engagement Analysis | Data Science Project (https://jingchenggu.github.io/EG-Social-Analysis/)

- Led a data analysis project for the Evil Geniuses Esports' social media team using **Python**, **Pandas**, and **Matplotlib**, processing over 3500 posts since January 2023 to identify peak engagement times and optimize content strategy across various platforms.
- Assessed the performance of various game titles and media types, highlighting the superior engagement of the DOTA2 account and photo media, while advising against the use of link media due to low engagement rates.
- Improved EG social media engagement rate by 60% through presenting easily comprehensible data-driven recommendations for the social media team, including strategic post scheduling, content focus on the game-specific accounts, and emphasis on specific media types.

Data Science Student Society | VP Internal (https://www.ds3ucsd.com/) — Jun 2023 - *Present*

- Spearheaded a dynamic interdisciplinary group of over 300 undergraduate data science students, conducting weekly meetings for all members, ensuring cohesive management and effective implementation of the organization's diverse initiatives.
- Initiated a mentorship program, pairing 30+ junior members with senior mentors, boosting member engagement by 50%.

TECHNICAL SKILLS

Languages: Python, Java, C/C++/C#, JavaScript, HTML, SQL, R

Frameworks / Libraries: Pandas, NumPy, PyTorch, TensorFlow, OpenCV, JUnit, Raspberry Pi Tools, Beautiful Soup, Scikit-Learn

Tools: Amazon Web Services (S3, Lambda, DynamoDB), Github, Git, Excel, Tableau, Power BI, IntelliJ, MongoDB, D3.js, Svelte