Longhao, Lin

lol008@ucsd.edu| (909) 244-2495 | linkedin.com/in/longhaolin08 | github.com/lh-008

SUMMARY

Aspiring Data Analyst with a strong foundation in Python, Data Visualization, and Data Wrangling. Currently pursuing a Bachelor's degree in Data Science, with a focus on applying statistical methods to uncover insights. Eager to leverage academic projects and adaptable learning skills to contribute effectively to data-driven decision-making.

EDUCATION

University of California, San Diego
Bachelor of Science, Data Science **(Expected) 06/2026**

Relevant Courseworks: Principle of Data Science | Programming and Data Structure for Data Science | Theoretical Foundation of Data Science | Practice and Application of Data Science

SKILLS

Programming Languages: Python | Java | HTML
Libraries/Packages: Pandas | Numpy | Sci-Kit Learn | Beautiful Soup | NLTK | Regex | Seaborn | Matplotlib
Tools and Technologies: Git | Github | API | Web Scraping | Regression Analysis | Statistical inference
Natural Languages: English (Fluent) | Mandarin (Fluent)

PROJECTS

Python Image Processing **March - 2024**

- Designed and implemented core image manipulation methods like negation, grayscale conversion, and brightness adjustment without relying on external libraries.
- Developed a machine learning model using Euclidean distance to classify images, incorporating robust voting mechanisms for label prediction.
- Engineered dynamic app structures using object-oriented programming to track user costs and provide exclusive features, such as advanced image tiling and sticker integration.
- Designed algorithms to efficiently handle complex image transformations, prioritizing scalability and runtime optimization.

N-Gram Language Model **November - 2024**

- Scraped and processed public domain books from Project Gutenberg using web scraping techniques and regular expression for data extraction and cleaning
- Tokenized large text efficiently with regular expression, ensuring proper handling of punctuation, whitespace, and special characters.
- Designed and implemented baseline language models, including Uniform and Unigram Language Models.
- Built an N-Gram Language Model from scratch, including generating N-Gram tokens, training the model, and calculating token probabilities with statistical accuracy.
- Engineered the model to generate coherent sentences under optimized runtime conditions.

League of Legend Dataset Analysis **December - 2024**

- Analyzed 150k + observations from professional League of Legends match data using Python (Pandas, Numpy) for data cleaning and exploratory data analysis.
- Create visualizations with Plotly, to uncover trends, patterns, and relationships between game objectives and match statistics.
- Applied hypothesis/permutation testing to effectively determine statistical significance and missing mechanism in the data.
- Engineered optimal features to develop a binary classifier that predicts game outcomes, achieving an accuracy of 97% and an F1-score of 97%
- Conducted fairness analysis to ensure the model performs equitably across different groups.