Garvey J. Li

(510)589-5163 | garvey.j.li@gmail.com |www.linkedin.com/in/garveyli/ | www.github.com/GarveyJLi

EDUCATION

University of California - San Diego Sept 2021 - June 2025

B.S. Data Science, Minor in Cognitive Science *Cum Laude* - 3.9 GPA

Relevant Coursework: Database Design/Architecture, Data Visualization, Representation Learning, Probabilistic Modeling and ML, Systems for Scalable Analytics, Reinforcement Learning

SKILLS

Tools: Python, Java, Regex, Bash, SQL, Github, Git, Spark, Databricks, Snowflake, Excel, Powerpoint

Python Packages: Pandas, Sklearn, Dask, PySpark, Pytorch, Keras, XGBoost, Alteryx, Matplotlib

WORK EXPERIENCE

Franklin Templeton - Investment Management Data Science Intern Jun 2024 - Aug 2024

(Python, PySpark, SQL, Databricks, Snowflake, Sklearn, Hyperopt, Powerpoint)

Conducted research on financial data to analyze changes in security performance indicators caused by major economic events such as the 2008 recession and the COVID-19 pandemic.

- Wrangled and cleaned financial data from various sources (e.g. CapIQ), and engineered new features that may have been indicative of security performance.
- Created a Random Forest Classifier using Bayesian hyperparameter optimization and a custom cross validation method to account for data leakage present in sequential finance data.
- Feature importance was analyzed using mean decrease in impurity and permutation importances.
- *Due to confidentiality agreements, specific results and business impact cannot be disclosed.

Glass Labs - Bioinformatics Research Assistant Mar 2024 - Present

(Python, Pytorch, HOMER, Bash)

- Researched new ways to encode genetic data to enable better use of machine learning in analysis.
- Utilized transformer models for motif detection and researched ways to interpret attention layers to find relationships between tokens (base-pairs).

PROJECTS

Dino-Nuggetology - DS3@UCSD DataHacks 2024 (1st Place Overall) Apr 2024

(Python, Pytorch, Scikit-Image, Scipy, OpenCV, PIL) https://devpost.com/software/dino-nuggetology

Created a model that classifies if a dinosaur shaped chicken nugget is normally shaped or malformed.

- Manually collected data by taking pictures of 228 unique dino-nuggets, and used edge detection to represent images as binary vectors.
- Made augmented versions of images to generate more data, resulting in ~10,000 data points.
- Trained a convolutional autoencoder to detect anomalies based on reconstruction accuracy.

GPT-2 Transformer Based Music Generation Jun 2025

(Python, Transformers, PyTorch) https://garveyjli.github.io/ragtime-generator

Trained and fine-tuned a GPT-2 LM Head Model to generate ragtime music with the REMI tokenizer

- Webscraped 425 ragtime genre MIDI files and conducted EDA to validate the files and their contents.
- Used the REMI tokenizer to encode and represent the music files as lists of tokens symbolizing notes, chords and time based events that occur throughout the music.
- Fine-tuned the GPT-2 model on the resulting tokens, using the CrossEntropyLoss metric.
- Resulting model could generate music that followed the general patterns of ragtime music.