Lily Yu

858-683-9768 | me@lily-pad.io | github.com/yu-lily

Education

University of California San Diego — Expected June 2023
Bachelor of Science in Data Science — *San Diego, CA*

- **GPA**: 3.86
- **Notable Coursework**: Database Systems, Recommender Systems, Robotics Perception & Navigation, Topological Data Analysis (Graduate level)

Technical Skills

Languages: Python, JavaScript, SQL (prior experience: Java, Lua, C++, R, Matlab)
Tools: Git, Jupyter, Docker, HTML/CSS, LaTeX, soldering iron
Libraries: numpy, pandas, scikit-learn, BeautifulSoup4, Selenium, matplotlib, networkx, gudhi
AWS: Cloud Development Kit (IaC), Lambda, API Gateways, S3, EC2, ECS, DynamoDB, Lightsail

Experience

Computational Topology Research Intern — June 2021 – Present
Wang Lab, UC San Diego

- Explored feasibility of geometric/topological modeling for time-varying materials science data
- Improved time complexity of graph skeleton approximation from $O(n^3)$ to $O(n^2 \log n)$ by applying state-of-the-art graph contraction algorithms

Data Engineering Assistant — Feb. 2021 – Present
CLIVAR and Carbon Hydrographic Data Office

- Modernized AWS infrastructure to take advantage of serverless frameworks (Lambda)
- Reduced runtime of batch file conversion job from 12 hours to 5 minutes by containerizing function and leveraging scalable on-demand compute (Lambda, ECS)
- Saved 50+ engineering hours over one month by creating a reusable system for parallelizing functions formalized with an Infrastructure as Code framework (AWS CDK)

Projects

Fantasy DotA 2 Toolkit | *Python* — October 2021

- Utilized to achieve highest 2021 fantasy league score in a UC-wide competition (and highest 0.1% of scores globally) with an ensemble model trained on historical and live data
- Used OpenCV and Tesseract to continuously ingest live data throughout duration of tournament

Computational Topology Music Classifier | *Python* — May 2021

- Final project for graduate-level course on Topological Data Analysis methods
- Improved classification accuracy by 8% using Persistent Homology and Hierarchical Clustering Trees
- Surpassed baselines using Hierarchical Clustering Trees for classification
- Final Report: https://yu-lily.github.io/tda-final/

DataHacks 2021 - LSTM Stock Prediction | *Python, Javscript, D3.js* — April 2021

- Placed 1st among 40+ teams, 150+ participants in data-centric hackathon
- Predicted S&P 500 values with contest-winning accuracy using multiple LSTM architectures
- Received the only unanimous perfect score from judges with a reactive website report to communicate methods
- Final Report: https://yu-lily.github.io/datahacks-2021-report/

DotA 2 Replay Archiver | *Python, AWS Lambda* — June 2020 – August 2020

- Deployed a scalable system that archives 3TB per month by executing queued web request actions in parallel
- Reduced AWS costs by 97% by creating a model to curate the highest-skill matches involving professional players