(202) 834-8536
Rosslyn, VA
ml1859@georgetown.edu

Eric Lim

Data Scientist / Software Engineer

Portfolio
GitHub
LinkedIn

Skills

Data Science	Python (pandas, numpy, scikit-learn, scipy, tensorflow, pytorch, nltk), SQL (PostgreSQL, MySQL, MongoDB), R (ggplot2, tidyverse)
Data Engineering	AWS, Docker, Kubernetes, Kubeflow, Jenkins, Gradle
Analytics	Tableau, D3, Plotly, matplotlilb, seaborn
Back-end	Java, C/C++, Solidity
Front-end	HTML, JS, CSS
Tools	Version Control (Git, SVN), Jira (Kanban, Scrum), Bitbucket, Artifactory

Professional Experience

Data Platform Engineer — **Jul 2023 — Present**
Freddie Mac — *McLean, VA (Hybrid)*

- Upgraded data platform infrastructure through AWS EKS, EC2, and CFTs to expedite data ingestion, data pipelining, and data warehousing
- Initiated onboarding with external vendors and performed cross-coordination new tech ingestion with internal teams
- Created video demos and user documentation on software and feature updates for ease of access and use for data science teams

Software Development Engineer — **May 2022 — Aug 2022**
Amazon — *Nashville, TN (Remote)*

- Improved business scalability and flexibility by revamping the Java back-end to allow API calls and DJS jobs to handle multiple data types
- Reduced AWS S3 bucket storage needs by 55% by enhancing prediction models to require a shorter time frame for generating similar forecasts
- Facilitated more targeted forecasting for different vendors by creating new data pipelines and migrating existing model data flows to new AWS databases
- Refined the onboarding process to accelerate ramp-up for future hires with detailed tech setup documentation

Projects

Data Visualization: Transportation Accidents in the U.S. — **May 2023**

- Led a visual analysis on the U.S. transportation sector with interactive data visualizations of railroad and airplane accidents

✤ **Techniques:** Data Mining, Data Wrangling, EDA

Statistical Analysis: Hans Niemann Cheating Scandal — **Dec 2022**

- Led a statistical analysis on a chess cheating scandal to provide insight on the validity of Magnus Carlsen's allegations against Hans Niemann
- Discovered relevant KPIs in determining a player's legitimacy through EDA and linear regression analysis that shaped the rest of the investigation

✤ **Techniques:** Data Mining, Data Wrangling, EDA, Shapiro-Wilk, Kruskal-Wallis, Linear Regression, Multicollinearity tests, Bootstrapping, T-tests

Machine Learning: Gender Equality in the U.S. — **Dec 2022**

- Analyzed gender equality in the U.S. by observing educational, occupational, and societal metrics with supervised and unsupervised machine learning models
- Gathered and cleaned structured and non-structured data from government sources and social media with Python and R scripts

✤ **Techniques:** Web Scraping, Data Mining, Data Wrangling, EDA, Naive Bayes, Decision Trees, Random Forests, SVM, Clustering, ARM, Networking

Media Playlist Simulator — **Dec 2021**

- Implemented a Spotify-like Flask app that stores users' log-in information and respective playlists of songs, movies, TV shows, and podcasts with Python and SQL

ATM Simulator — **Dec 2021**

- Developed the functions of an ATM, such as handling users' banking details, depositing/withdrawing money, and sending spending reports with Java and Python

Education

Master of Science in Data Science and Analytics, *Georgetown University*, GPA: 4.0 — Aug 2022 - May 2024
Bachelor of Science in Computer Science and Statistics, *Georgetown University*, GPA: 3.6 — Aug 2019 - May 2023