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

Eric Lim

Data Scientist / Software Engineer

Portfolio
GitHub
LinkedIn

Skills

Data Science	Python, SQL, R, Spark, Tableau
Cloud	AWS, Microsoft Azure
Data Engineering	Docker, Kubernetes, Kubeflow, Jenkins, Gradle
Software Engineering	Java, C/C++, Solidity, HTML, JS, CSS, Version Control

Professional Experience

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

- Established an automated monthly report generation system of KPIs for supported data science teams and platforms
- Upgraded data platform infrastructure via AWS to expedite data ingestion, data pipelining, and data warehousing workflows
- Spearheaded the creation of custom Docker images used by various data science teams to enable faster onboarding of resources

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

Big Data: Politics on Reddit based on the Economy **Dec 2023**

- Conducted analysis on political subreddits to identify trends in user posts, comments, and interactions in parallel with changes in the U.S. economy

✤ **Techniques:** Data Mining, Data Wrangling, Spark, EDA, NLP, Sentiment Analysis, Linear Regression, Logistic Regression, Random Forests

MLOps: Credit Card Fraud Detection **Dec 2023**

- Coordinated the development and deployment of a machine learning app that detects fraudulent credit card transactions with an accuracy of 99%.

✤ **Techniques:** MLOps, EDA, Logistic Regression, k-Nearest Neighbors, Decision Trees, SVM, Naive Bayes, Random Forests, Bagging, Gradient Boost, XGBoost, Stacked Models

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
- Engineered statistical models to determine a player's legitimacy

✤ **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**

- Launched 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

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