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

# Eric Lim

Data Scientist / Software Engineer

Portfolio
GitHub
LinkedIn

## Skills

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

## Professional Experience

**Data Engineer** **Jul 2023 — Present**
*Freddie Mac* *McLean, VA (Remote)*
- Spearheaded the creation and usage of custom Docker images, lowering turnaround time for onboarding three teams of over 50 data scientists and streamlining data science workflows
- Guided and oversaw upgrades to data infrastructure and software used by data scientists, speeding up model training from over 48 man-hours to just a few minutes
- Outlined and initiated the creation of a new automated monthly KPI report generation system, leveraging in-house tools and reducing reliance on external vendors

**Software Development Engineer** **May 2022 — Aug 2022**
*Amazon* *Nashville, TN (Remote)*
- Discovered inefficiencies in prediction models and revamped them to require a shorter time frame for generating similar forecasts, reducing data storage costs by 55%
- Identified and addressed future scalability bottlenecks by splitting one large use case into three separate use cases in the codebase
- Overhauled existing forecasting data pipelines and upgraded data flow to accommodate these new use cases

## 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**
- Authored 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