

MINSUH (ERIC) LIM

TBD, PA ZIPCODE

📞 202-834-8536 ✉ ml1859@georgetown.edu [🌐 LinkedIn](#) [🐙 Github](#) [📁 Portfolio](#)

Education

Georgetown University

Master of Science, Data Science and Analytics. GPA: 4.0 / 4.0

Bachelor of Science, Computer Science and Statistics. GPA: 3.6 / 4.0

Washington, DC

Aug 2022 – May 2024

Aug 2019 – May 2023

Experience

GSK

Data Scientist

Aug 2024 – Present

Philadelphia, PA (Hybrid)

- Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras tellus dolor, tristique eu orci sed, dictum faucibus magna. Phasellus facilisis luctus malesuada.

Freddie Mac

Data Engineer

Jul 2023 – Jul 2024

McLean, VA (Remote)

- Spearheaded creating and using 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.

Amazon

Software Development Engineer

May 2022 – Aug 2022

Nashville, TN (Remote)

- Discovered inefficiencies in prediction models and revamped them to require a shorter time frame for generating 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.

Technical Skills

Data Science: Python, SQL, R, Tableau, Spark, NLP, Deep Learning, PyTorch, TensorFlow, Keras

Data Engineering: Docker, Kubernetes, Kubeflow, Jenkins

Software Engineering: Java, C/C++, Linux, HTML, JS, CSS, Version Control

Cloud: AWS, Microsoft Azure

Projects

Deep Learning: Subject-level Similarity Art Finder | Python, Streamlit

May 2024

- Deployed a Streamlit app that processes a user's image and returns art pieces from DeepLake that are similar using ResNet50 and YOLOv5.

Machine Learning: StarCraft II Player Performance Prediction | Python, R

May 2024

- Analyzed in-game metrics to predict a player's rank with 40% accuracy and 80% ROC-AUC score using ANOVA Tests, Pairwise T-Tests, Linear Regression, Logistic Regression, SVMs, XGBoost, and Random Forests.

NLP: Medical Abstract Text Classification | Python

May 2024

- Developed a Python package for medical text data using Word2Vec and Doc2Vec to calculate Cosine and Jaccard similarities, and RNNs, XGBoost, and Naive Bayes to perform classification.

Big Data: Politics on Reddit based on the Economy | Python, SQL, Spark, AWS, Microsoft Azure

Dec 2023

- Analyzed political Subreddits to identify trends in user activity following changes in the U.S. economy using Azure (Machine Learning), AWS (S3, Redshift, Athena, SageMaker, Lambda), PySpark, and SparkSQL.

MLOps: Credit Card Fraud Detection | Python, Microsoft Azure, Heroku

Dec 2023

- Built a Heroku app that detects fraudulent credit card transactions with an accuracy of 99% using a stacked model consisting of XGBoost, Bagging, Logistic Regression, and Random Forests.

Data Visualization: Transportation Accidents in the U.S. | Python, R, Tableau

May 2023

- Performed dashboarding of U.S. railroad and airplane accidents using Plotly, Shiny, and Tableau.

Statistical Analysis: Hans Niemann Cheating Scandal | Python, R

Dec 2022

- Evaluated a chess cheating scandal using Linear Regression, Bootstrapping, Shapiro-Wilk Test, Kruskal-Wallis Test, Multicollinearity Tests, and T-Tests.