Minji Lee

Allston, MA 02134 | lee.minji9521@gmail.com | 857-316-7741
LinkedIn- https://www.linkedin.com/in/minji-lee-7b2126120 | Github- https://github.com/mjlee2121

WORK EXPERIENCE

Wood Mackenzie, Full Stack Software Engineer I — February 2022 – May 2024

Energy Management Team

- Developed and implemented RESTful API endpoints to handle complex data flows and built Swagger docs
- Wrote Python scripts to calculate energy flow from original ISOs to individual customers
- Created unit tests to identify missing data points or fault forecasts daily and implemented an automated email notification system which identified 20% of data being inaccurate
- Refactored static National Grid webpage scrapers from BeautifulSoup to Selenium and built a pipeline to generate csv files for analysis, reduced loading time from 30 minutes to 7 minutes

Metals and Mining Team

- Developed frontend components of metals and mining dashboard to organize and display data reports using TypeScript, React, and Mapbox
- Engineered a data pipeline to publish data to an external platform API utilizing AWS SageMaker, S3, and Step Functions
- Communicated with stakeholders including project managers, analysts, and UX/UI designers and built presentation slides to confirm the design directions and technical implementations
- Contributed to ETL processes including data preprocessing, log management in Datalake, and deployment to AWS Athena and Redshift
- Designed AWS Athena schema to store 50GB of data to S3
- Maintained account security by assigning different policies and entitlements to accounts using microservices console

Regeneron Pharmaceuticals Inc, Data Science Co-op (Intern) — January 2021 – August 2021

Antibody Sequence Informatics

- Preprocessed and cleaned 100GB antibody sequence data (.fasta) from PDB with Python
- Applied language models to build a model that can predict proteins involved in antibody-antigen binding

Gait Analysis with Force Plates and Wearable Insoles

- Preprocessed the given proprietary data including normalization, standardization, and interpolation with Python
- Built models using TPOT, automated machine learning toolkit, that is trained on golden standard vertical ground reaction force (vGRF) data collected from force plates and tested on vGRF collected from wearables to predict injured subjects
- Achieved F1 score of 0.88, proved that our hypothesis is correct
- The finding of this research has been published on MedRxiv under the title: 'Machine learning analysis of a digital insole versus clinical standard gait assessments for digital endpoint development'

EDUCATION

Northeastern University, MA — Spring 2019 – Fall 2021
Master of Science in Data Science
Related Courses: Supervised Machine Learning, Unsupervised Machine Learning, Natural Language Processing, Big Data for Cities

Bridgewater State University, MA — Spring 2015 – Spring 2018
Bachelor of Science in Computer Science, Minor in Statistics
Related Courses: Data Structure & Algorithms, Introduction Database Systems, Operating Systems, Analysis of Algorithms, Organization of Program Languages, Object-Oriented Software Engineering, Senior Design & Development, Statistical Methods I & II, Regression Analysis, Design & Analysis Experiments

Kyungpook National University, Daegu, South Korea — Spring 2013 – Spring 2016
Five semesters of course work in Chinese Language and Literature; and Computer Science
Related Courses: Fundamentals in C/C++, Introduction to Computer Science & Engineering

TECHNICAL SKILLS

Languages & Platforms: Python, Java, C, SQL, R, Git, JavaScript, TypeScript, HTML, CSS, React, Docker, AWS SageMaker, GCP, Jenkins, Django, Celery, Postgres
Hardware: Arduino
Database & Query Languages: AWS Athena, Redshift, S3, GraphQL, OData

PERSONAL PROJECTS

New York Times

- Constructed a news website mimicking New York Times using newsapi and JavaScript
- Implemented features such as search bar, pagination, and categorization

City Explorations

- Analyzed gentrification indicators on Building Permits data, hosted by The Boston Area Research Institute, with R
- Visualized Geographic Information using *ggmap, sf,* and *ggplot2*
- Published work weekly on WordPress blog