

Mengmeng Liu

410-585-7689 • 07302 Jersey City NJ • mengmengliu24@gmail.com • [Github](#) • [Linkedin](#)

EDUCATION

JOHNS HOPKINS UNIVERSITY

Master of Science in Engineering, Data Science

Teaching Assistant – EN.553.636 Introduction to Data Science (Grad Level)

Baltimore, MD

Aug 2021 – Dec 2022

Feb 2022 – Dec 2022

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

Chengdu, China

Bachelor of Science, Math and Applied Math

Sept 2016 – July 2020

PROFESSIONAL EXPERIENCE

Data Engineer

Feb 2023 - Current

Healthcare Risk Advisors

New York City, NY

- Developed real-time machine learning services for a new web portal, including a data quality scoring and alerting service for loss run files. Achieved 95% accuracy and resulting in significant improvement in the quality of processed data. Utilized Azure MLflow and Databricks serving endpoints to manage and optimize.
- Created a Python package for Azure Databricks Feature Store and MLflow API to enable seamless model development and deployment. Worked with data science team to integrate models into production environment.
- Collaborated closely with cross-functional teams to understand and complete their data requirements, answer technical questions about data models, and develop customized Power BI reports and DAX formulas.
- Managed end-to-end data pipeline operations for the data compliance tool, ensuring consistent data quality and availability
- Worked as a member of a development team to create a web application using React and JavaScript. Contributed to project planning and agile development processes with Microsoft DevOps.

Data Engineer Intern

Jun 2022 – Dec 2022

Healthcare Risk Advisors

New York City, NY

- Built a data compliance tool using PySpark, Spark SQL and power BI to empower the business team to effectively verify data quality and compliance with around 70 compliance rules for over 8,000 human-coded healthcare cases. This tool enabled the team to quickly and accurately identify data errors and efficiently correct them, resulting in a 50% reduction in data errors.

Data Analyst Intern

Dec 2020 – May 2021

Yimian by Ascential

Shenzhen, China

- Developed and refactored auto data cleaning pipeline for O2O e-commerce project using Python and Impala SQL, increased data and pipeline reuse rate by 30%; Conduct O2O data asset project sharing meeting
- Researched and trained unsupervised item image classification model by Pytorch with accuracy around 75%; Deployed through designing RESTful API by Flask, Unicorn and docker, reduced repeated image cleaning and labelling process by 80%

PROJECTS

Affiliation Name Disambiguation on IEEE Content Database

Sept 2022 – Dec 2022

Developer

Baltimore, MD

- Built and deployed an HAC model using NLP word2vec embeddings to address affiliation name ambiguity in IEEE's data lake with 5.8M+ research articles. Achieved 96% in-cluster accuracy on representative dataset.
- Constructed end to end AWS data cleaning solution from data collection to model scoring with AWS S3 and EC2.
- Optimized the pipeline's execution time and input data size by 30% by incorporating Joblib and multiprocessing.

University Admission Prediction Google Cloud Application

Nov 2021 – Dec 2021

Designer and Developer

Baltimore, MD

- Built up data modeling pipeline and trained machine Learning models with accuracy rate around 80%.
- Designed web application using Flask, HTML and CSS; Deployed the model through Google Cloud App Engine.

SPECIALIZED SKILLS

- Coding: advanced SQL, Python (Scikit-Learn, Pytorch, Numpy, Pandas, SciPy), C, Matlab
- Software & Skills: Databricks, Git, Linux, VS code, Pycharm, Jupyter Notebook, Latex, Power BI