

MIN (MIA) SHI

408-698-1020 ◇ miashil406@gmail.com ◇ [Personal Portfolio](#) ◇ <https://www.linkedin.com/in/min-mia-shi/>

EDUCATION

The University of Texas at Dallas

M.S. in Business Analytics (Data Science & Data Engineer Track)

M.S. in Social Data Analytics and Research

May 2024

GPA: 3.95/4.0

GPA: 4.0/4.0

SKILLS

Programming & Tools: Python, Java, Git, R, SQL, SAS, Stata, Tableau, Power BI, Alteryx

Frameworks & Tools: Django, AWS, Hbase, Redis, Memcached, RabbitMQ, Celery, Amazon S3

Database & Big Data: MySQL, PostgreSQL, Hadoop, Sqoop, Hive, Impala, Pig, Spark

Certificates: Graduate Certificate in Applied Machine Learning, AWS Certified Cloud Practitioner

WORK EXPERIENCES

Research Assistant

The University of Texas at Dallas

May 2020 - Present

Richardson, TX

Took responsibility for data manipulation and model building for 10+ global health and policy analytics projects.

- Developed and automated data collection processes for over 10 projects using Python, R, and SQL.
- Engineered efficient, reliable data pipelines in a Hadoop ecosystem, reducing processing time by 40%.
- Built and implemented statistical models for data analysis, including time-series and multi-variable regression.

AI Chatbot Student Consultant

Working for Onyx CenterSource through The University of Texas at Dallas

Aug. 2023 - Dec. 2023

Dallas, TX

Led the creation of an AI-driven chatbot, enhancing customer engagement through advanced NLP techniques.

- Utilized NLP to parse data and MySQL to query a database with over 10 million entries.
- Enhanced response efficiency by 25% and achieved 99% accuracy in predictions using the XGBoost model.
- Boosted user engagement by 15%, increasing customer satisfaction and strengthening the company's image.

Technology Data Analyst

Lucion Technology Corp., Ltd.

Jul. 2017 - Aug. 2017

China

Served as a Technology Analyst Intern responsible for data management, data visualization, and business analysis.

- Improved the efficiency of data extraction by 40% through data optimization in MySQL.
- Employed Microsoft Visio to visualize intricate network structures and aided in product comprehension.
- Produced Business Intelligence reports, offering insights based on extensive user data and competitor analysis.

PROJECTS

Twitter Clone: High-throughput Social Media Backend — *Python, Django*

Jan. 2024 - Present

Working on a three-month solo project developing a social media platform's backend using HBase, MySQL, and Redis with Django framework in Python.

- Maximize query efficiency by storing objects with HBase, MySQL & Amazon S3 based on query complexity.
- Address N+1 slow query issues by implementing Redis caching and denormalization.
- Integrate Celery and RabbitMQ to establish asynchronous workers with varying priority levels.
- Implement a push model for distributing news feeds to followers efficiently.
- Optimize memory and resource allocation using recursive small batches of asynchronous tasks.

Kaggle Plant Pathology Competition: Leveraging Deep Learning CNNs

Nov. 2023 - Dec. 2023

Implemented deep learning models using Python and PyTorch to enhance disease identification accuracy in crops.

- Utilized transfer learning on CNNs with 13042 images in 12 categories, enhancing identification accuracy.
- Conducted image transformation, including rotation, flipping, zooming, and noise injections to augment data.
- Fine-tuned ConvNext DL CNN models and achieve 87% accuracy, securing a Top 3 ranking in the competition.

Big Data Risk Analysis and Data Visualization for a Trucking Company

Aug. 2022 - Dec. 2022

Engineered data visualization dashboards using Tableau, linked to Hadoop, for business risk analysis.

- Processed and analyzed geospatial data with Hadoop, Hive, and Spark, reducing processing time by 40%.
- Developed Tableau visualizations linked to Hadoop and built interactive dashboards for business analysis.
- Conducted linear regression and multivariate analysis, contributing to predictive accuracy by 15%.