MIN (MIA) SHI

minmiashi@gmail.com ◇ Personal Portfolio ◇ https://www.linkedin.com/in/min-mia-shi/

WORK EXPERIENCES

Data Scientist & Data Engineer Jun. 2024 - Present
The Sunwater Institute *North Bethesda, MD / Remote*
Developed scalable ETL pipelines and real-time data workflows, implemented speech-to-text and speaker identification with 99% accuracy, and optimized AI/ML models for speech processing and big data using AWS.

- Developed real-time data processing workflows using AWS S3, Glue, and Lambda, optimizing performance and ensuring high efficiency. *Tools: AWS S3, Glue, Lambda, Python (Boto3)*
- Automated the speech-to-text and speaker identification process using AWS Transcribe, Computer Vision, and Deep Learning Models using PyTorch, achieving an accuracy rate above 90%, and contributing to speech processing initiatives. *Tools: PyTorch, CNN, Transfer Learning, AWS Transcribe & Textract*
- Engineered and maintained scalable data applications, incorporating data quality checks and implementing automated CI/CD pipelines for seamless deployments. *Tools: Streamlit, Python, GitHub, AWS EC2, Jenkins*
- Collaborated with cross-functional teams, including engineers, researchers, and stakeholders, to design and implement AI-based data models for projects such as speech identification.
- Authored documentation for data models and workflows, fostering better collaboration and transparency.

Data Analyst May 2020 - May 2024
The University of Texas at Dallas *Richardson, TX*
Took responsibility for data analysis for 10+ global health/policy projects using advanced statistical models.

- Managed data collection in diverse methods including Qualtrics surveys and web scraping using R and Python.
- Developed 20+ robust statistical models (multi-variable and fixed-effect regression, difference-in-difference, time-series), utilized ML models and NLP skills to support correlation and causal inference in research.
- Supervised a team of five junior assistants, ensuring timely project completion and publication.

PROJECTS

AI Chatbot Project Aug. 2023 - Dec. 2023
Led the development of an AI-driven chatbot for Onyx CenterSource, improving customer engagement through advanced NLP techniques.

- Leveraged NLP and MySQL for analyzing and querying an extensive database containing over 10 million entries.
- Improved response efficiency by 25% and achieved 99% accuracy using the XGBoost model.
- Enhanced user engagement, boosting customer satisfaction and strengthening the company's brand image.

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.

- Applied transfer learning on CNNs with 13,042 images across 12 categories, significantly improving accuracy.
- Performed image augmentation techniques (rotation, flipping, zooming, noise injection) to enhance data.
- Fine-tuned ConvNext DL models, achieving 86.8% accuracy and 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 significant processing time.
- 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%.

EDUCATION

Ph.D. in Political Science	The University of Texas at Dallas	Dec. 2024 (***GPA: 3.95/4***)
M.S. in Business Analytics	The University of Texas at Dallas	May 2024 (***GPA: 3.95/4***)

SKILLS

Programming & Tools: Python, SQL, AWS Web Services (S3, Glue, Lambda, and Boto3), Streamlit, R
AI & ML Modeling: PyTorch, Deep Learning, Machine Learning, NLP, Speech Processing, Real-Time Model Deployment, Speech-to-Text, Speaker Identification
Database & Big Data: SQL Server, MySQL, PostgreSQL, AWS RDS, Hadoop, Spark, Hive, Impala, Sqoop, Pig
Certificates: Certificate in Applied Machine Learning, AWS Certified Cloud Practitioner, Google Analytics