

MIN (MIA) SHI

Teaching Assistant, School of Economic, Political & Policy Sciences at UTD
Cecil H. Green Hall | 800 West Campbell Road, GR 3.314 | Richardson, TX 75080-3021
Min.Shi@utdallas.edu ◇ minshimia@gmail.com
[UTD Profile](#) ◇ [Personal Portfolio](#)

Education

The University of Texas at Dallas Ph.D. Candidate in Political Science – Quantitative Statistical Modeling Focused	Aug. 2019 – Dec. 2024 <i>GPA: 3.95/4.0</i>
The University of Texas at Dallas M.S. in Social Data Analytics and Research	Aug. 2021 – Aug. 2024 <i>GPA: 3.95/4.0</i>
The University of Texas at Dallas M.S. in Business Analytics (Data Science & Data Engineer Track)	Aug. 2022 – May 2024 <i>GPA: 4.0/4.0</i>
The University of Texas at Dallas Graduate Certificate in Applied Machine Learning	Aug. 2022 – May 2023 <i>GPA: 4.0/4.0</i>
The University of Texas at Dallas M.A. in Political Science	Aug. 2019 – May 2022 <i>GPA: 3.95/4.0</i>
Shandong University M.L. in International Politics	Sept. 2016 – Jun. 2019 <i>GPA: 88.78/100</i>
Daito Bunka University Exchange Student in Political Science	Sept. 2017 – Aug. 2018
Shandong University B.A. in Japanese	Sept. 2012 – Jun. 2016 <i>GPA: 87.37/100</i>

Scholarships

Keith Lankford Tayer Fellowship	2024
John Forrest Kain Scholarship	2023
Government and Political Science Scholarship	2022

Work Experiences

The Sunwater Institute <i>Data Scientist</i> ↔ North Bethesda, MD / Remote	Jun. 2024 – Present
<ul style="list-style-type: none">• Summary: Managed data collection, developed data pipeline, built analytics dashboards, and provide infrastructure support.• Implemented web scraping solutions to extract data from websites, storing over 1 million records in databases.• Created ETL process for ingesting data using AWS S3 and Glue, boosting data processing efficiency by 40%.• Automated speech-to-text and speaker identification using AWS Transcribe, achieving over 99% accuracy.• Developed and maintained interactive Shiny dashboards, ensuring scalability and usability.• Managing supporting infrastructure, including database and system administration.	
The University of Texas at Dallas <i>Data Analyst Research Assistant</i> ↔ Richardson, TX / Part-time	May 2020 – May 2024
<ul style="list-style-type: none">• Summary: Took responsibility for data manipulation and model building for 10+ global health and policy analytics projects.• 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) combined ML models and NLP skills to support correlation and causal inference in research.• Led a team of five junior assistants, ensuring collaboration and timely project completion and publication.	

Onyx CenterSource *Data Science Student Consultant*

Aug. 2023 – Dec. 2023

↪ Dallas, TX / Consultant

- **Summary:** Led the creation of an AI-driven chatbot, enhancing customer engagement through advanced NLP techniques.
- Employed NLP and MySQL for analyzing and querying an extensive database containing over 10 million entries.
- Achieved 25% improvement in response efficiency and provided 99% accurate predictions using XGBoost model.
- Contributed to a 15% rise in user engagement, increasing customer satisfaction and bolstering company's image.

Lucion Technology Corp., Ltd. *Marketing Data Analyst*

July 2017 – Aug. 2017

↪ Jinan, CN / Intern

- **Summary:** Served as a Data 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 (BI) reports, offering insights based on user structures and competitor analysis.

Data Science & ML & Backend Projects

Twitter Clone: High-throughput Social Media Backend

May 2024 - Present

- **Summary:** Optimizing a social media backend using technologies like HBase, MySQL, and Redis.
- Maximizing query efficiency by storing objects with HBase, MySQL, and Amazon S3 based on query complexity.
- Addressing N+1 slow query issues by implementing Redis caching and denormalization.
- Integrating Celery and RabbitMQ to establish asynchronous workers with varying priority levels.
- Implementing a push model for distributing news feeds to followers efficiently.
- Optimizing memory and resource allocation using recursive small batches of asynchronous tasks.

US Top 4 Airlines Financial Performance Analytics

Jan. 2024 - May 2024

- **Summary:** Analyzed 20 years of airline data, identified strategic trends and turning points, and recommended business models.
- Analyzed financial data from a 20-year dataset of over 10,000 rows, covering net income, revenue, and expenses across the US airline industry. This deep dive provided insights into long-term financial trends and shifts.
- Conducted financial performance analytics for the top 4 airlines, identifying key turning points related to major events, alliances, and partnerships over the period.
- Assessed operational trends and competitive positioning of each airline, deriving specific business model recommendations based on a two-decade comparison with competitors.

Kaggle Plant Pathology Competition: Leveraging Deep Learning CNNs

Nov. 2023 - Dec. 2023

- **Summary:** 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 disease identification accuracy.
- Conducted image transformation, including rotation, flipping, zooming, and noise injections to augment data.
- Fine-tuned ConvNext DL CNN models and achieve 86.8% accuracy, securing a Top 3 ranking in the competition.

Forecasting Stock Prices Through NLP Examination of Newspaper Articles

May 2023 - Dec. 2023

- **Summary:** Developed automated web scraping, applied NLP techniques to analyze WSJ articles, and improved S&P 500 prediction accuracy.
- Developed automated web scraping for 7,000+ WSJ articles, increasing data acquisition efficiency by 30%.
- Employed various vectorizers for WSJ article analysis, such as Tfidf Vectorizer, n-grams Count Vectorizer, etc.
- Utilized Naïve Bayes and Random Forest models, enhancing S&P 500 prediction accuracy by 12%.

Analysis of the Effect of COVID-19 on US Trade and US Firms

May 2023 - Jul. 2023

- **Summary:** Built regression and machine learning models for causal analysis, and presented findings at the 2023 Applied Data Science International Conference, earning recognition.
- Synthesized data and created fixed-effect regression models to identify correlations and causal mechanisms.

- Developed and Implemented machine learning and deep learning models to conduct counterfactual analysis.
- Presented research at the 2023 Applied Data Science International Conference to 200 professionals, receiving recognition for clarity and actionable insights.

Analytical Insights and Marketing Strategy Guidance for a Food Company

Feb. 2023 - May 2023

- **Summary:** Cleaned 1.3 million data records, built interactive Tableau dashboards, and improved forecasting accuracy by 15%.
- Handled data cleaning over 1.3 million raw data records using Python, ensuring data quality and accuracy.
- Developed interactive dashboards in Tableau, enhancing data accessibility and supporting business analytics.
- Employed SAS to construct regression and time series models, leading to a 15% increase in forecasting accuracy.

Big Data Risk Analysis and Data Visualization for a Trucking Company

Aug. 2022 - Nov. 2022

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

Payroll Management System Database Design via MySQL

Jun. 2022 - Aug. 2022

- **Summary:** Designed and implemented a payroll management database in MySQL, leading a team of five; developed automated functions, procedures, and triggers, and optimized ETL processes and queries.
- Led a group of five in conducting business requirements analysis and designing a payroll management database with MySQL consisting of 13 tables.
- Created stored functions, procedures, and triggers to calculate employees' payroll per two weeks, fill in new employee information, and send PTO reminders automatically.
- Performed extract-transform-load, data cleaning, and query optimization.

Modeling U.S.-China Trade War's Effect on US Firms using ML and Time Series

Jan. 2022 - May 2022

- **Summary:** Analyzed the impact of the US-China trade war on MNCs using ML, sentiment analysis, and GARCH time series models.
- A project aimed at exploring how the US-China trade war affects Multinational Corporations (MNCs) through an ML content analysis of policy changes and a time series GARCH modeling approach using stock data.
- Utilized Pandas, NumPy, Matplotlib & Seaborn in data cleaning, visualization, and transformation.
- Leveraged sentiment analysis to explore how the US frame 2018 US-China trade war
- Applied regression analysis in exploring the causal mechanism between trade war and S&P 500 revenues.
- Built machine learning (ML) models in predicting the profound influence of the trade war on US firms.
- Used time-series GRACH models to evaluate MNCs' revenue & volatility quantified via stock data in Stata.
- Presented at 2022 International Society for Data Science and Analytics Conference.

Content Analysis of News Coverage about US-China Trade War

Aug. 2022 - May 2022

- **Summary:** Analyzed how news organizations framed the 2018 US-China trade war, using machine learning and time-series analysis on over 500 articles to model sentiment trends.
- Led an analysis on how news organizations frame the 2018 US-China trade war during the 2018-2022 period.
- Leveraged machine learning skills such as top modeling and sentiment analysis to explore a collection of over 500 news articles.
- Implemented time-series analysis and chi-squared test in modeling sentiments change tendencies among news coverage.

COVID-19 Worldwide Cases Synchronous Dashboard using Tableau

Dec. 2021 - Jan. 2022

- **Summary:** Designed and developed an interactive Tableau dashboard to analyze COVID-19 severity worldwide, uncovering key factors influencing the pandemic's impact across countries.
- Designed a synchronous Tableau dashboard with advanced interactive functions to explore the COVID-19 severity.
- Built a Tableau story to dig into the factors affecting the severity of COVID-19 by country and found out the deep connection between multiple aspects of factors with COVID-19 severity.

Selected Course Work

Data Science	Data Management	Data Modeling
Deep Learning	Big Data	Predictive Analytics for Data Science
Natural Language Processing	Cloud Computing Fundamentals	Modeling for Business Analytics
Causal Analytics and A/B Testing	Database Fundations for BA	Regression and Multivariate Analysis
Programming for Data Science	Information Management	Applied Data Analytics with Python
ML for Socio-Eco and Geo Data	Data Collection	Applied Regression
Content Analysis using ML	Data Visualization	Introduction to Quantitative Methods
OOP in Python	Digital Consulting Project	Social Science Research Methodology
Data Structure & Algorithm	Practical Practicum Project	Prescriptive Analytics

Technical Skills

Programming	Python, R, SQL, Java, Stata, SAS
Tools	Alteryx, Tableau, Jupyter Notebook, Amazon Web Services, Excel Charts, R Shiny, \LaTeX & \TeX
Database	MySQL, PostgreSQL, SQL Server, Mango DB, Amazon RDS
Big Data	Hadoop, Sqoop, Hive, Impala, Pig, Spark
Automation	Alteryx, Appian, Acceleq, Uiopath
Certificates	Graduate Certificate in Applied Machine Learning at UTD, Google Data Analytics, AWS Certified Cloud Practitioner, Alteryx Designer Core Certificate, Appian Certified Associate Developer, ACCELQ Automation Engineer
Languages	English, Chinese, Japanese