Mirza Nayeem Ahmed

 github | linkedin | 🌐 website | ✉ ahmed.mir@northeastern.edu | 📱 +1.508.231.6555

SUMMARY

Data Engineer transitioning to Computational Social Science, leveraging extensive analytical skills in both corporate and academic environments. Currently focused on advancing expertise in computational methods and fieldwork in preparation for Ph.D. program applications for Fall 2025. Based in Boston, Massachusetts, seeking research opportunities and open to relocation to further academic and career pursuits in social science research.

EDUCATION

Northeastern University (Boston, MA)

Bachelor of Science, **Data Science**   Sep 2019 - May 2023

Minor: Entrepreneurship & Design Engineering

GPA: 3.66/4.0

Honors: Dean's List, Distinction

Relevant Courses: Discrete Structures, Biostatistics, Database Design, Natural Language Processing, Object-Oriented Design, Network Science, Web Development, Human-Computer Interaction, Machine Learning

RESEARCH

University of South Florida   Nov 2023 - Present

Research Associate, Department of Computer Science and Engineering

Developing simulation models for analyzing social networks and testing social science theories at the Computational Social Science Lab under Professor Raiyan Abdul Baten. The research focuses on the influence of social interactions on belief rigidity and creativity, employing agent-based models to explore complex social dynamics. Key tasks include evaluating relevant literature to identify research gaps, conducting rigorous hypothesis testing through advanced statistical methods, and performing computational simulations using Python and R. The ongoing project aims to contribute to the field through the publication of influential findings and the development of coding frameworks that facilitate future research within this domain.

Research Areas

Computational Social Science, Network Science, Data Science, Organizational Behavior, Human Dynamics

Research Interests

Behavioral Analytics, Fraud Detection, Risk Modeling, Machine Learning in Social Predictions, Affective Computing, Large-Scale Data Architecture, Digital Sociology, Multilevel Modeling

WORK EXPERIENCE

ANIQUE, LLC (Remote)   Oct 2023 - Present

Data Consultant

– Lead R&D initiatives to develop and implement an advanced, scalable data architecture for a dynamic e-commerce product catalog, ensuring secure and reliable cloud-based database operations

– Establish strategic partnerships with external companies to acquire valuable data sources, facilitating data enrichment for informed business decisions

– Implement AI algorithms for product attribution, search, and filtering to enhance UX and increase customer engagement

Condé Nast (New York, NY)   Jun 2022 - Aug 2022

Data Analyst

– Built scalable ETL pipelines that support consumer events of the 16+ brands such as Vogue, GQ, WIRED, and The New Yorker for 2M+ customers using Apache Spark, Airflow, S3, and Databricks Delta Tables; enabling real-time data processing, providing visibility into critical metrics, and reducing data latency by 15%

– Applied scripting expertise to securely transfer sensitive information (database credentials, tokens, passwords, certificates, and API keys) to support vendor migration, enhancing data security and mitigating unauthorized access risks

– Leveraged Terraform to automate the provisioning and management of Condé Nast's global data platform, reducing infrastructure deployment time by 20% and providing a robust foundation for data processing and analytics

– Researched and developed a proof of concept for an in-house analytics tool to monitor social media engagement, securing buy-in from senior management (CTO, VP of Marketing, and SVP of Audience Development)

McKinsey & Company (Boston, MA)   Jul 2021 - Dec 2021

Data Engineer

– Developed and implemented an automated recovery processing system in Domo to validate data, analyze usage, and calculate monthly chargeback for 15+ services, reducing manual steps and saving over 50% in ETL time

– Created and deployed a consolidated report visualizer for the finance and operation teams, centralizing critical data, providing real-time updates and ensuring data security and privacy compliance for over 200 employees

– Implemented AI-driven capabilities and leveraged external data sources to optimize a knowledge-aggregation tool for the Client Capabilities Hub, resulting in a 35% reduction in business analysts' information access time while ensuring data synchronization, CI/CD, and facilitating swift diagnosis and resolution

Northeastern University (Boston, MA)   Aug 2020 - Dec 2022

Teaching Assistant

– Courses taught across different semesters: CS3200 Database Design, GE1110 Engineering Design, GE2500 Design Analysis, DS2000 Programming with Data, DS4300 Large Scale Information Storage and Retrieval

– Facilitated 400+ hours of weekly office hours and helped 700+ students with course work; designed rubrics, and graded 1000+ quizzes and assignments across the semesters

– Mentored students in research and design projects, fostering academic growth and innovative thinking

– Guided students in internship preparation to enhance their professional readiness

PROJECTS

Prison Book Project Data Visualization

Python (seaborn, altair, bokeh), Tableau, Javascript (d3), MongoDB

• Designed and implemented complex data visualizations to provide insights into various aspects of the non-profit organization, including prison restrictions, donation patterns, and volunteer activities

• Collaborated with the board of directors to identify operational inefficiencies within the program and streamline the book collection and delivery process

• Automated data cleaning and validation processes for diverse and unstructured data sets, ensuring sustainable project performance with minimal need for future intervention

Parks and Recreation Tweet Generator

Python (keras, numpy, tensorflow, tweepy), MySQL

- Developed an end-to-end solution building and deploying a recursive neural network model and a statistical model to generate novel sentences using dialogues from a TV show, offering a unique and efficient approach to content creation
- Integrated the Twitter API to automate the uploading of tweets generated by the sentence generator at specified time intervals, alleviating the need for manual tweet creation and scheduling

Gene Association and Drug Interaction Dashboard

Python (pandas, plotly dash, sklearn), Neo4j

- Deployed a dashboard that connects gene association data with drug interaction information, allowing users to find out which drugs can potentially interact adversely based on their genetic profile, facilitating informed medication decisions
- Wrangled and mapped out relationships within the datasets, ensuring efficient storage and retrieval of data
- Leveraged machine learning for clustering and predictive modeling, offering personalized insights and recommendations

Skills

Data Engineering	Apache Spark, Airflow, dbt, Databricks, Snowflake, Alteryx
Cloud & DevOps	Terraform, Amazon Web Services, Docker
Data Science & Analytics	Python, R, Scala, Domo, Altair, Tableau
Databases	MySQL, Oracle SQL, MongoDB, Redis, Neo4j
Network Science	NetworkX, NetLogo
Web Development	Figma, HTML, CSS, Java, Javascript
Embedded Systems	Arduino, MATLAB
Engineering Design	AutoCAD, Solidworks
Technical Writing	LaTeX
Project Management	JIRA, Confluence

Languages

English	fluent
Bengali	bilingual proficiency
Hindi	limited working proficiency

References

Dr. Raiyan Abdul Baten

Assistant Professor, College of Engineering

University of South Florida

rbaten@usf.edu

Dr. John Rachlin

Associate Teaching Professor, Khoury College of Computer Sciences

Northeastern University

j.rachlin@northeastern.edu

Archan Bhatt

Director of Engineering & Associate Partner

McKinsey & Company

archan_bhatt@mckinsey.com