Madhurima Nath

+1 (540) 235-2449 | madhurima.nath@icloud.com LinkedIn: madhurimanath | GitHub: madhurima-nath

Data scientist with PhD in computational physics, specializing in machine learning, NLP, and data engineering. Experience delivering end-to-end solutions across consulting, spanning data pipeline development, predictive analytics, and GenAI applications.

Skills

Programming & Data Processing: Python, SQL, PySpark, R, Bash || Databricks, Snowflake, CI/CD || AWS, Azure, GCP
Machine Learning & GenAI: MLflow, LangChain, TensorFlow || RAG, Prompt Engineering, Multi-Agent Frameworks
Project Management & Collaboration: Agile/Scrum (JIRA), cross-functional team leadership, stakeholder communication

Professional Experience

Slalom, Inc., New York, New York, USA
Senior Data Scientist Sep 2024 – Present
Data Scientist Jul 2021 – Aug 2024
Associate Data Scientist Jan 2020 – Jun 2021

- ***GenAI Projects***

  Product Catalogue Discovery Assistant (2 months) || US Fortune 500 Financial Services Client || Lead Data Scientist

  - Designed and developed a RAG system on Databricks integrating Llama with LangChain to enable internal advisors to search and retrieve product information from internal catalogues.
  - Configured MLflow for experiment tracking and model versioning, with Unity Catalog managing the vector database for product embeddings, coordinating with compliance teams to establish data governance controls.

  Clinical Data Support Q&A Chatbot (6 weeks) || Healthcare Technology Startup || Data Scientist

  - Developed Q&A assistant using Anthropic Claude on AWS Bedrock to translate clustering algorithm outputs into natural language summaries for healthcare providers and researchers.
  - Applied prompt engineering to combine clustering results with structured patient data including symptoms, health conditions, and drug classifications from JSON files.

- ***Key Data Science & Data Engineering Projects***

  NLP for Safety Analytics (5 months) || US Fortune 500 Energy Utility Client || Lead Data Scientist

  - Developed and deployed end-to-end NLP pipeline on Azure Databricks to process five years of safety incident reports, implementing topic modeling and sentiment analysis to extract safety themes from unstructured incident narratives.
  - Deployed production solution with automated weekly reports and interactive dashboards, receiving executive recognition and wide organizational adoption for providing actionable intelligence on safety policy and resource allocation.

  Data Lake Implementation on Snowflake (12 months) || Nonprofit Healthcare Organization

  - Maintained and upgraded Snowflake data lake processing 2M daily JSON bundles of statewide patient health information, building task orchestrations and parameterized stored procedures ensuring compliance with data protection regulations.
  - Refactored legacy monolithic JavaScript into modular Python and SQL components, improving maintainability and code quality.

Research Experience

Post-doctoral Research Assistant, Virginia Tech, Virginia, USA **Feb 2019 – Dec 2019**

- Developed hybrid computational approach combining Monte Carlo simulations with perturbative expansions to improve estimation of Moore-Shannon network reliability on graphs, addressing an NP-hard computational problem. GitHub repo
- Applied computational methods to analyse international trade network dynamics from the United Nations (UN) Comtrade database, to identify crucial communities for preventing global pestilence distribution. Publication

Education

Ph.D., Physics (Computational), Virginia Tech, Virginia, USA Dec 2018
M.S., Physics, Virginia Tech, Virginia, USA May 2017
M.Sc., Physics, Indian Institute of Technology Delhi, India May 2012
B.Sc. (Hons.), Physics, University of Calcutta, India May 2010

Outreach

Peer reviewer for NeurIPS, ICML, Physical Review E
Industry Ambassador, Women in Network Science Society (2022–2025)
Multiple peer-reviewed publications in network science, statistical physics, and conference presentations at SIAM, APS, NetSci.

Click for Extended CV