Madhurima Nath

madhurima.nath@icloud.com || LinkedIn: madhurimanath Google Scholar || GitHub: madhurima-nath

Computational physicist and machine learning researcher specialising in artificial intelligence (AI) & machine learning (ML) solutions, statistical physics, and complex systems. Experience spans development of novel network analysis methods and implementation of large-scale machine learning systems, with particular focus on probabilistic modeling and scalable data analytics. Proven track record of developing production-ready ML solutions while leading cross-functional teams in engagements.

Education & Experience

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

Post-doctoral Research Assistant, Virginia Tech, VA — Feb 2019 – Dec 2019

Ph.D., Physics, Virginia Tech, Blacksburg, VA — Dec 2018
- *Research Focus*: Complex systems analysis, Statistical physics, Network science, Community detection
- *Dissertation*: Application of Network Reliability to Analyse Diffusive Processes on Graph Dynamical Systems
- *US Patent* (US20210286859A1): System, method and computer readable medium for sensitivity of dynamical systems to interaction network topology

M.S., Physics, Virginia Tech, Blacksburg, VA — May 2017

M.Sc., Physics, Indian Institute of Technology Delhi, New Delhi, India — May 2012
- *Thesis*: Study of Cold Atomic Condensates by Atomic Photon Interactions
- *Award*: Best Master of Science Thesis 2011-2012

B.Sc. (Hons.), Physics, University of Calcutta, Kolkata, India — May 2010
- *Minors*: Mathematics and Chemistry

Machine Learning Research & Development (2020–Present)

Generative AI & Chatbot Proof of Concepts (POCs)

- **Lead Data Scientist || Product Discovery Chatbot POC || Fortune 500 Financial Services Client**
  Architected a product discovery system on Databricks integrating Llama[1] with the LangChain agent framework, implementing MLflow pipelines. Developed a scalable solution in 2 months that achieved measurable improvements in information retrieval.
- **Data Scientist || Interactive Q&A Chatbot || Healthcare Tech Startup**
  Developed a Q&A system using Anthropic Claude model on AWS[2] Bedrock, designed to extract insights from complex bio-pharma and med-tech datasets, in 6 weeks. Enhanced user interface design to improve accessibility and user satisfaction, resulting in a scalable solution for healthcare sector decision-making.
- **Data Scientist || Knowledge-Based Chatbot for Sales || Fortune 500 Equipment Rental Firm** Engineered and deployed a knowledge-based chatbot using Retrieval Augmented Generation (RAG) with OpenAI GPT[3] and LangChain, improving document search efficiency and reducing response times to inquiries by 22%, delivering the first digital innovation initiative for the client.

Data Science Projects

- **Lead Data Scientist || NLP[4] for Safety Analytics || Fortune 500 Energy & Utility Client**
  Led the development & deployment of advanced NLP[4] pipelines on Azure Databricks, utilizing sentiment analysis and topic modeling to extract critical safety-related insights. Empowered the executive safety committee by providing data-driven analysis that refined organisational policies, directly contributing to improved employee safety and measurable reductions in incident frequency and severity.
- **Lead Data Scientist || NLP[4] for e-Discovery Document Evaluation || Legal Services Client**
  Architected an e-discovery document classification system using Word2Vec document embeddings and SVM[5] classifiers on AWS[2] SageMaker, implementing an automated retraining pipeline that incrementally updates models based on human-reviewed samples. Demonstrated potential for reducing manual review efforts and costs through automated document classification and feedback-driven model improvements.
- **Data Scientist || NLP[4] for Global Procurement || Fortune 500 Consumer Packaged Goods Company**
  Developed text classification pipeline to harmonise procurement product data into a standardised template using RandomForest classifier on Azure Machine Learning platform. Implemented MLOps best practices to incorporate addition of new products, updates to data labels and model re-training. Accelerated the data foundation roadmap by over a year and built scalable infrastructure design that improved data consistency and performance.

[1] *Llama:Large Language Model Meta AI*, [2] *AWS: Amazon Web Services*, [3] *GPT: Generative pre-trained transformer*, [4] *NLP: Natural Language Processing*, [5] *SVM: Support Vector Machines*

Data Engineering Experience

Data Migration to Salesforce || Nonprofit Organisation

- Successfully migrated 1TB of historical data to Salesforce through meticulously designed SQL stored procedures and optimised ELTL[6] pipelines in Azure Data Factory, ensuring comprehensive data integrity and seamless integration.
- Implemented advanced validation techniques that significantly reduced migration errors and enhanced overall data quality for enterprise-level data management.

Data Lake Implementation on Snowflake || Nonprofit Organisation

- Maintenance and upgrade of data pipelines to load patient health and healthcare related data through parameterised SQL stored procedures in Snowflake Data Lake, ensuring comprehensive data integrity and seamless integration.
- Validated large-scale data quality checks and data transformation rules across multiple data resources providing and sharing patient health information.

Network Predictive Maintenance Solution || Telecom Client

- Engineered a statistical radio frequency metrics analysis system with geographical clustering to proactively identify faulty infrastructure before customers face a major service disruption.
- Implemented automated SQL procedures processing 90-day network performance data, enabling targeted preventive maintenance validated across multiple regions.
- Projected $8M+ savings in annual operational costs through reductions of ∼30K customer support calls and ∼6K service visits, based on initial analysis later refined for regional variations.

Academic Research Experience

Post-doctoral Research Assistant, Virginia Tech, VA, (2019)

- Improved the complexity of evaluating Moore-Shannon network reliability on a graph (an NP[7]-hard problem) by combining Monte Carlo simulations with weak- and strong-coupling perturbative methods. Technical paper
- Implemented statistical approaches for community detection in large-scale weighted directed networks, achieving significant improvements over traditional methods.
- 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. Community detection paper

Graduate Research Assistant, Virginia Tech, (2014-2018)

- Developed new approaches to Moore-Shannon network reliability formalism using Bernstein basis functions, Monte-Carlo simulations and statistical perturbative methods. GitHub repo
- Applied Moore-Shannon network reliability to predict final global states of graph dynamical systems, analyzing how interactions between individual node states and their connections determine final outcomes in practical applications. PhD Dissertation
- Simulated real-world epidemic outbreak and intervention scenarios on National Longitudinal Study of Adolescent to Adult Health dataset using network reliability, providing policymakers with improved estimates. Existing methods show about 50% over-estimation in number of infections. Epidemic paper
- Identified vulnerabilities within global food trade networks and accurately forecasted the impact of mitigating contagion conditions with approximately 96% precision to understand and mitigate the spread of pests. Food trade paper
- Developed an algorithm using concepts of Moore–Shannon network reliability to estimate the energy states of an interacting magnetic system, reducing the time complexity from $O(n^2)$ to $O(n)$. Ising model paper

Technical Expertise

Computational & Data Processing

- Programming: Python, SQL, PySpark, R, Bash
- Distributed Computing: Apache Spark, PySpark, Databricks || Cloud Platforms: AWS[2], Azure, GCP[8],
- Version Control & CI/CD[9]: GitLab, GitHub, Azure DevOps, BitBucket, Docker (familiar)

Machine Learning Analytics

- ML Frameworks: MLflow, LangChain, TensorFlow (familiar) || Data Tools: Delta Lake, Delta Lakehouse
- Areas: Statistical modeling, NLP[4], LLMs[10] (Llama[1], Anthropic Claude, OpenAI GPT[3])

[1] *Llama:Large Language Model Meta AI*, [2] *AWS: Amazon Web Services*, [3] *GPT: Generative pre-trained transformer*, [4] *NLP: Natural Language Processing*, [6] *ELTL: Extract, Load, Transform, Load*, [7] *NP: non-deterministic polynomial-time hardness*, [8] *GCP: Google Cloud Platform*, [9] *CI/CD: : Continuous Integration & Continuous Development*, [10] *LLM: Large Language Model*

Publications

- Mishra, R., Eubank, S., **Nath, M.**, Amundsen, M., and Adiga, A. (2022, November). Community Detection Using Moore-Shannon Network Reliability: Application to Food Networks, In International Conference on Complex Networks and Their Applications (pp. 271-282). Cham: Springer International Publishing.
- Eubank, S., **Nath, M.**, Ren, Y., and Adiga, A. (2022). Perturbative methods for mostly monotonic probabilistic satisfiability problems. arXiv preprint arXiv:2206.03550.
- **Nath, M.,** et. al. (2019). Using network reliability to understand international food trade dynamics. In Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7 (pp. 524-535). Springer International Publishing.
- **Nath, M.**, Ren Y., and Eubank, S. (2019). An approach to structural analysis using Moore-Shannon network reliability. In Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7 (pp. 537-549). Springer International Publishing.
- **Nath, M.**, Ren, Y., Khorramzadeh, Y., and Eubank, S. (2018). Determining whether a class of random graphs is consistent with an observed contact network. Journal of theoretical biology. 440, 121-132.
- **Nath, M.,** and Eubank, S. (2018). Model selection for sequential designs in discrete finite systems using Bernstein kernels. arXiv preprint arXiv:1807.06661.
- Ren, Y., Eubank, S., and **Nath, M.**. (2016). From network reliability to the Ising model: A parallel scheme for estimating the joint density of states. Physical Review E, 94(4), 042125.
- **M. Nath**, et. al. (2015). A two-parameter method to characterize the network reliability for diffusive processes. In Complex Networks VI: Proceedings of the 6th Workshop on Complex Networks CompleNet 2015 (pp. 139-148). Springer International Publishing.
- Agarwala, A., **Nath, M.**, Lugani, J., Thyagarajan, K., and Ghosh, G. (2012). Fock-space exploration by angle resolved transmission through a quantum diffraction grating of cold atoms in an optical lattice. Physical Review A, 85(6), 063606.

Conference Presentations

- **Invited Speaker**: Network reliability: a generic tool to explore diffusive processes on interacting systems, NASA Prebiotic Chemistry and Early Earth Environments (PCE3) Virtual Workshop 2022 Nano- to Cosmic- Studies of Complex Systems, University of Wisconsin Madison, Madison, WI, 20 Oct, 2022.
- Perturbative methods for estimating relative contributions to network reliability, SIAM (Society for Industrial and Applied Mathematics) Workshop on Network Science, Virtual Workshop, 13-15 Sep, 2022.
- Statistical mechanical applications of graph dynamical systems, Condensed Matter Seminar, Department of Physics, Virginia Tech, Blacksburg, VA, 30 Oct, 2017.
- Determining whether a particular contact network is consistent with a network model, 1st North American Social Networks Conference of the International Network for Social Network Analysis, Washington DC, 26-30 Jul, 2017.
- Network reliability: A novel measure to study the effects of network topology on the diffusive dynamics, Symposium for the Society of Young Network Scientists, NetSci 2017, Indianapolis, IN, 19-23 Jun, 2017.
- Network reliability: A measure to study diffusive dynamics on networks, Center for Soft Matter and Biological Physics Symposium 2017, Virginia Tech, Blacksburg, VA, 17-18 May 2017.
- Effects of network structure on propagation of infectious diseases, 33rd Annual Graduate Student Assembly Symposium and Exposition, Virginia Tech, Blacksburg, VA, 29 Mar 2017.
- Renormalisation group approaches for dynamics on irregular networks, American Physical Society (APS) March Meeting 2017, New Orleans, LA, 13-17 Mar, 2017.
- Diffusive dynamics on a network, Southeastern Section of the American Physical Society (SESAPS) Conference 2016, Charlottesville, VA, 9-12 Nov, 2016.
- Effects of network structure on epidemic modeling, Biocomplexity Institute Symposium 2016, Virginia Tech, Blacksburg, VA, 1 Nov, 2016.
- A two-parameter method to characterise the network reliability for diffusive processes, CompleNet 2015, New York City, NY, 25-27 Mar, 2015.
- Four-parameter characterisation of network reliability and analysis of critical point phenomenology, APS (American Physical Society) March Meeting 2015, San Antonio, TX, 2-6 Mar, 2015.

Professional Service & Academic Engagement

Technical Review & Service

- Reviewer, NeurIPS: Machine Learning and Physical Sciences (2022, 2024)
- Reviewer, ICML[9]: Synergy of Scientific and Machine Learning Modeling (2023)
- Reviewer, Physical Review E (2022-Present)
- Industry Ambassador, Women in Network Science Society (2022-Present)
- Organiser of Women in Network Science Networks 2021 Conference (2021)

Teaching & Mentorship

- Managed & mentored summer interns in advanced analytics and ML solutions, Slalom (2021-Present).
- Instructor of Record, Department of Physics, Virginia Tech (2018)
  Developed and delivered comprehensive physics curriculum for a 3-credit introductory physics course covering Waves, Acoustics & Optics to $\sim$ 90 undergraduate students.
- Graduate Teaching Assistant, Department of Physics, Virginia Tech (2013-2015)
  - Supervised laboratory sessions for introductory physics courses (Kinematics & Electro-magnetism) and provided individualised student support.
  - Recognised as a top performing physics instructor for creating an effective learning environment.

Invited Online Lectures

- **Speaker**: Insights into Data Engineering (Parts 1 & 2), Women Who Code San Francisco chapter, Jul 2021.
  YouTube: Part 1 YouTube: Part 2
- **Speaker**: Exploring NLP[4] Fuzzy Matching Algorithms, Women Who Code, Apr 2021. YouTube, Medium article
- **Panelist**: Statistics in Data Science Discussion, Women Who Code Statistics in Data Science Workshop Series, Mar 2021. YouTube
- **Speaker**: Regression & Predictions, Women Who Code Data Science: Statistics Workshop Series - Statistics in Data Science, Feb 2021. YouTube, Medium article

[4] *NLP: Natural Language Processing*, [9] *ICML: International Conference on Machine Learning*,