

# Madhurima Nath

madhurimanath21@gmail.com || LinkedIn: madhurimanath

Google Scholar || GitHub: madhurima-nath

Computational physicist and machine learning researcher specializing 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

<b>Senior Data Scientist</b> , Slalom, Inc., New York, NY	Sep 2024 – Present
<b>Data Scientist</b> , Slalom, Inc., New York, NY	Jul 2021 – Aug 2024
<b>Associate Data Scientist</b> , Slalom, Inc., New York, NY	Jan 2020 – Jun 2021
<b>Post-doctoral Research Assistant</b> , Virginia Tech, VA	Feb 2019 – Dec 2019
<b>Ph.D., Physics</b> , Virginia Tech, Blacksburg, VA	Dec 2018
<ul style="list-style-type: none"><li><i>Dissertation</i>: Application of Network Reliability to Analyze Diffusive Processes on Graph Dynamical Systems</li><li><i>Research Focus</i>: Complex systems analysis, Statistical physics, Network science</li><li><i>US Patent</i> (US20210286859A1): System, method and computer readable medium for sensitivity of dynamical systems to interaction network topology</li></ul>	
<b>M.S., Physics</b> , Virginia Tech, Blacksburg, VA	May 2017
<b>M.Sc., Physics</b> , Indian Institute of Technology Delhi, New Delhi, India	May 2012
<ul style="list-style-type: none"><li><i>Thesis</i>: Study of Cold Atomic Condensates by Atomic Photon Interactions</li><li><i>Award</i>: Best Master of Science Thesis 2011-2012</li></ul>	
<b>B.Sc. (Hons.)</b> , Physics, University of Calcutta, Kolkata, India	May 2010
<ul style="list-style-type: none"><li><i>Minors</i>: Mathematics and Chemistry</li></ul>	

## MACHINE LEARNING RESEARCH & DEVELOPMENT (2020–PRESENT)

- Generative AI & Chatbot Proof of Concepts (POCs)**
  - Lead Data Scientist || Product Discovery Chatbot POC || Financial Services Client  
Architected a product discovery system on Databricks integrating Llama<sup>1</sup> with the LangChain agent framework, implementing MLflow pipelines. Developed a scalable solution in 2 months achieving 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<sup>2</sup> Bedrock, designed to extract insights from complex biopharma and medtech 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 || Equipment Rental Firm  
Engineered and deployed a knowledge-based chatbot using Retrieval Augmented Generation (RAG) with OpenAI GPT<sup>3</sup> and LangChain, improving document search efficiency and reducing inquiry response times by 22%, delivering the first digital innovation initiative for the client.
- Key Data Science Projects**
  - Lead Data Scientist || NLP<sup>4</sup> for Safety Analytics || Energy & Utility Client  
Led the development & deployment of advanced NLP<sup>4</sup> 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 organizational policies, directly contributing to enhanced employee safety and measurable reductions in incident frequency and severity.
  - Lead Data Scientist || NLP<sup>4</sup> for e-Discovery Document Evaluation || Legal Services Client  
Architected an e-discovery document classification system using Word2Vec document embeddings and SVM<sup>5</sup> classifiers on AWS<sup>2</sup> 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 || Syndicated Retail Data Analysis || Consumer Packaged Goods Company  
Implemented fuzzy matching algorithms to consolidate syndicated retail data from multiple vendors (Nielsen, IRI<sup>6</sup>, and Skupos) into a centralized Azure Databricks platform. Eliminated manual data reconciliation processes, achieving a 38% efficiency improvement in product comparison and analysis. Leveraged sophisticated data integration techniques to transform complex, disparate data sources into a unified, actionable intelligence repository.

<sup>1</sup>Llama: Large Language Model Meta AI, <sup>2</sup>AWS: Amazon Web Services, <sup>3</sup>GPT: Generative pre-trained transformer, <sup>4</sup>NLP: Natural Language Processing, <sup>5</sup>SVM: Support Vector Machines, <sup>6</sup>IRI: Information Resources, Inc.

## ACADEMIC RESEARCH EXPERIENCE

---

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

- Formulated novel algorithms combining Monte Carlo simulations with perturbative methods for NP<sup>7</sup>-hard problems in complex systems analysis.
- Implemented statistical approaches for community detection in large-scale weighted directed networks, achieving significant improvements over traditional methods.
- Applied computational methods to analyze international trade network dynamics from the United Nations (UN) Comtrade database, to identify crucial communities for preventing global pestilence distribution.

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

- Developed new approaches to network reliability using Moore-Shannon formalism. [GitHub repo](#)
- Created efficient Monte Carlo methods to simulate real-world epidemic outbreak scenarios, providing policymakers with improved estimates.
- 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.
- 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  $O(n)$ .

## DATA ENGINEERING EXPERIENCE

---

### Data Migration to Salesforce || Nonprofit Organization

- Successfully migrated 1TB of historical data to Salesforce through meticulously designed SQL stored procedures and optimized ETL<sup>8</sup> 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.
- Demonstrated expertise in large-scale data migration, transformation, and validation across cloud platforms.

## PROFESSIONAL SERVICE & ACADEMIC ENGAGEMENT

---

### Technical Review & Service

- Reviewer, NeurIPS: Machine Learning and Physical Sciences (2022, 2024)
- Reviewer, ICML<sup>9</sup>: Synergy of Scientific and Machine Learning Modeling (2023)
- Reviewer, Physical Review E (2022-Present)
- Industry Ambassador, Women in Network Science Society (2022-Present)
- Organizer 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 ~ 90 undergraduate students.
  - Supervised laboratory sessions for introductory physics courses (Kinematics & Electro-magnetism) and provided individualized student support.
  - Recognized as a top-performing physics instructor for creating 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<sup>4</sup> Fuzzy Matching Algorithms, 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](#)

---

<sup>4</sup>NLP: Natural Language Processing, <sup>7</sup>NP: non-deterministic polynomial-time hardness, <sup>8</sup>ETL: Extract, Transform, Load, <sup>9</sup>ICML: International Conference on Machine Learning,

## 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. (*equal contribution as first author*).

## 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, Oct. 20, 2022.
- Perturbative methods for estimating relative contributions to network reliability, SIAM (Society for Industrial and Applied Mathematics) Workshop on Network Science, Virtual Workshop, Sep 13-15, 2022.
- Statistical mechanical applications of graph dynamical systems, Condensed Matter Seminar, Department of Physics, Virginia Tech, Blacksburg, VA, Oct. 30, 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, Jul. 26-30, 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, Jun. 19-23, 2017.
- Network reliability: A measure to study diffusive dynamics on networks, Center for Soft Matter and Biological Physics Symposium 2017, Virginia Tech, Blacksburg, VA, May 17-18, 2017.
- Effects of network structure on propagation of infectious diseases, 33rd Annual Graduate Student Assembly Symposium and Exposition, Virginia Tech, Blacksburg, VA, Mar. 29, 2017.
- Renormalization group approaches for dynamics on irregular networks, APS (American Physical Society) March Meeting 2017, New Orleans, LA, Mar. 13-17, 2017.
- Diffusive dynamics on a network, SESAPS (Southeastern Section of the American Physical Society) Conference 2016, Charlottesville, VA, Nov. 9-12, 2016.
- Effects of network structure on epidemic modeling, Biocomplexity Institute Symposium 2016, Virginia Tech, Blacksburg, VA, Nov. 1, 2016.
- A two-parameter method to characterize the network reliability for diffusive processes, CompleNet 2015, New York City, NY, Mar. 25-27, 2015.
- Four-parameter characterization of network reliability and analysis of critical point phenomenology, APS (American Physical Society) March Meeting 2015, San Antonio, TX, Mar. 2-6, 2015.

## TECHNICAL EXPERTISE

---

### Computational & Data Processing

- Programming: Python, SQL, PySpark, R, Bash
- Distributed Computing: Apache Spark, PySpark, Databricks || Cloud Platforms: AWS<sup>2</sup>, Azure, GCP<sup>10</sup>,
- Version Control & CI/CD<sup>11</sup>: 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<sup>4</sup>, LLMs<sup>12</sup> (Llama<sup>1</sup>, Anthropic Claude, OpenAI GPT<sup>3</sup>), Statistical modeling

### Project Management & Collaboration

- Agile methodology
- Cross-functional team leadership
- Stakeholder communication

---

<sup>1</sup>Llama: Large Language Model Meta AI, <sup>2</sup>AWS: Amazon Web Services, <sup>3</sup>GPT: Generative pre-trained transformer, <sup>4</sup>NLP: Natural Language Processing, <sup>10</sup>GCP: Google Cloud Platform, <sup>11</sup>CI/CD: : Continuous Integration & Continuous Development, <sup>12</sup>LLM: Large Language Model