Valmik Nahata

New York City, NY | (914) 584-8003 | valmik.nahata@gmail.com | valmiknahata

Education

University of California, San Diego Sep. 2024 – Present

Bachelor of Science in Data Science

- Data Structures & Algorithms, Quantitative Methods in Business, Introduction to Research Methods & Statistical Analysis, Microeconomics & Macroeconomics, Multivariable Calculus

Experience

Dartmouth-Hitchcock Medical Center Jun. 2024 – Feb. 2025

Machine Learning Research Intern

- Developed memory-efficient RAG system using SLMs for pathology report interpretation, reducing memory usage by 93%+ compared to traditional LLMs while maintaining retrieval accuracy across 9,500+ TCGA reports
- Implemented fusion search architecture integrating BM25 semantic similarity with FAISS quantized vector search to retrieve top-15 most relevant documents using HuggingFace transformers and all-MiniLM-L6-v2 embeddings
- Co-authored research manuscript and designed accompanying conference poster and slide deck presentation

Tree-Plenish Jun. 2022 – Sep. 2022

Data Automation Intern

- Developed Python automation pipeline using pandas and Google Sheets API with AWS-hosted SQL database to streamline financial tracking across 300+ partner schools, 3,200+ volunteers and 50,000+ sapling distributions
- Executed market research on New England community colleges and presented strategic outreach framework to executive team, outlining pathway to expand environmental programs to 10+ new institutional partners

Projects

The Early Economic Impacts of Transformative AI Apr. 2025 – May 2025

1st Place – Apart Research Economics of Transformative AI Sprint

- Co-developed novel temporal coherence framework for AI automation using GPT-4.1-mini to estimate effective time across 450+ O*NET tasks, identifying 8-hour coherence threshold for 80%+ automation
- Conducted literature review and authored discussion sections statistically modeling 2024-2026 automation timeline projections, identifying regulatory constraints, human premium effects, and cost-implementation factors

Milwaukee Bucks Fan Engagement Prediction System Feb. 2025 – Mar. 2025

3rd Place – Milwaukee Bucks & Modine Manufacturing Business Analytics Hackathon

- Directed 5-person team developing dual Random Forest models in Python using scikit-learn with Matplotlib and Seaborn visualizations, achieving 81%+ classification accuracy across propensity and churn analyses
- Presented winning solution to franchise executives featuring 3 strategic recommendations including, custom referral programs, IP crossover events, and Figma-prototyped mobile wayfinding app based on model insights

A Statistical Analysis of Crab Pulsar Giant Pulse Rates Feb. 2023 – Jul. 2024

Co-authored Publication – West Virginia University

- Implemented seasonal and solar proximity analysis to examine influences on 24,000+ Crab Pulsar giant pulses across 461-day study period in Python using NumPy and Astropy on shared JupyterHub infrastructure
- Conducted 1.55GHz L-band radio observations of Crab Pulsar using Green Bank Observatory's 20m telescope
- Identified irregular giant pulses from supernova SN2023ixf, enabling potential follow-up studies

Cover Edge-Based Triangle Counting Jan. 2023 – May 2024

Co-authored Manuscript – New Jersey Institute of Technology

- Integrated 22 sequential and 11 parallel triangle counting algorithms in C++ and collaboratively executed comprehensive benchmarking across 12 real-world SNAP and 12 synthetic Graph500 RMAT datasets
- Standardized 15 algorithms into formal pseudocode with LaTeX, ensuring reproducibility in publication

Digital Asset Algorithmic Trading System Aug. 2022 – Jan. 2023

Independent Project

- Generated $6,000+ in profit using algorithmic trading strategies for Team Fortress 2 and Counter-Strike 2 digital assets
- Developed Python automation scripts leveraging Backpack.tf and Marketplace.tf APIs to execute 2,400+ strategic transactions with real-time market analysis and price optimization

Technical Skills

Languages: Python, R, C++, JavaScript, HTML & CSS, Java, SQL, MATLAB, LaTeX

Frameworks & Libraries: pandas, Matplotlib, scikit-learn, NumPy, PyTorch, TensorFlow, Seaborn, React

Developer Tools: VS Code, Jupyter Notebook, RStudio, Git, AWS, Firebase, Figma, Excel

Interests: SLMs & LLMs, RAG, NLP, Data Visualization, Statistical Modeling, System Design