

ASHABARI MAJUMDAR

ashabarimajumdar@gmail.com · [LinkedIn](#) · [Github](#) · [Website](#)

PROFESSIONAL EXPERIENCE

University of Notre Dame

Graduate Researcher

South Bend, IN

Jul 2017 – May 2025

- Designed and deployed a **scalable end-to-end ETL pipeline** to ingest, process, and analyze **5+ TB** of nuclear physics data; applied **anomaly detection, statistical modeling, and signal processing** on peak-derived features using **Python and SQL**, enabling accurate identification and classification of key signal peaks.
- Built and automated a scalable **pattern recognition pipeline** on high performance computing (**HPC**) systems, applying **feature engineering** with **Python** and **Bash (Linux)** scripting, and generating **1,000+ data visualizations** for internal and external collaboration.
- Parallelized large-scale nuclear simulations in a **distributed environment using MPI and job arrays**, reducing processing time by **40%**.

Argonne National Laboratory

Lemont, IL

Data Science Trainee

Oct 2024 – Dec 2024

- Developed a healthcare cost prediction pipeline with **Python, Pandas, scikit-learn, Plotly, and GitHub**; implemented **data engineering (ingestion, validation, feature engineering)**, Linear Regression and **Random Forest Quantile Regression**.
- Built an image-based skin fold anomaly detection tool using computer vision techniques, including **image preprocessing and a convolutional neural network (CNN)** for images and **XGBoost** for tabular data, trained on labeled data.
- Applied **Large Language Model (LLM)** embeddings to unstructured healthcare claim texts, improving efficiency with dimensionality reduction while preserving accuracy.

Lawrence Livermore National Laboratory

Livermore, CA

Summer Intern

May 2023 – Aug 2023

- Integrated and parallelized detector simulations in **high performance computing (HPC)** system using **C++** language, improving performance across scientific workflows.

PROJECTS

- Built an end to end ML workflow in Python that cleaned real hospital datasets, trained and evaluated survival models, and added a lightweight GenAI retrieval augmented (RAG) explanation module to generate patient risk forecasts, clear explanations, and early risk alerts.
- Created a supply chain optimization model that forecasts demand, predicts failures and improves allocation decisions, reducing unplanned shortages and supporting better inventory planning.

SKILLS

- Applications: Fine tuning LLMs (Hugging Face), NLP with transformers and embeddings, supervised and unsupervised learning, predictive modeling, XGBoost, anomaly detection, time series forecasting, feature engineering, dimensionality reduction, supply chain analytics (demand forecasting, inventory modeling, failure analysis, vendor risk), linear and integer optimization, data validation.
- Tools: Python (PyTorch, Scikit-learn, MPI, Matplotlib, Pandas, SciPy, NumPy, PuLP), C++, SQL, Version Control/ Git, CI/CD, Shell scripting (Bash), APIs, Tableau, Google Colab, Jupyter, VS Code.
- Cloud & Data: Cloud Computing (AWS, Azure, GCP), Databricks, HPC clusters, distributed computing, batch processing, parallel pipelines.

EDUCATION

Ph.D. in Physics (GPA: 3.8 / 4), University of Notre Dame

Jul 2017 – May 2025 | Notre Dame, IN

M.Sc. in Physics, Indian Institute of Technology Kanpur

Aug 2014 – Jun 2017 | Kanpur, India

AWARDS AND LEADERSHIP

AWARDS: Graduate Professional Development Awardee (2022, 2019, 2018); National Science Foundation USA Leadership Advancing Socially Engaged Research Scholar (2019); American Institute of Physics Science Communication Grant Winner (2019); Bachelors and Master's tuition fee support and stipend Govt. of India for being in top 1% among ~15 million students (2011-2016).

LEADERSHIP: University Association of Women in Science (Executive Board); American Association of University Women (AAUW) Campus Initiative (Communication Director).