

Mohamed Azarudeen Allabaksh

New York, NY | +1 (631) 352-6030 | azarudeen.allabaksh@gmail.com | GitHub | LinkedIn

Education

Master of Science in Computer Science
Stony Brook University, SUNY

Aug 2024 – May 2026
Stony Brook, NY

Technical Skills

Languages: Python, SQL, Shell/Bash

ML & Research: PyTorch, TensorFlow, Scikit-learn, Time-Series Forecasting, Scientific Computing, Data Assimilation, Experiment Design

Data Engineering: Pandas, NumPy, Xarray, ETL Pipelines, Data Cleaning, Multi-Source Data Fusion

Geospatial & Time-Series: Geospatial Data Processing, NetCDF, Time-Series Modeling

Databases: PostgreSQL, DynamoDB, Snowflake, Redis

Cloud & DevOps: AWS (SageMaker, S3, EMR, Lambda, CloudWatch), Docker, GitHub Actions, Git

Experience

Software Engineer I
Stony Brook University

Aug 2024 – Jan 2026
New York, NY

- Developed **Python PyTorch models** incorporating real-time observations from multiple sensor sources into forecasting pipelines, processing **8K–12K** daily data assimilation events with **99.4%** pipeline reliability across **4 data source types**.
- Built **data cleaning and fusion pipelines** for large, messy datasets with inconsistent formats and resolutions, combining **Pandas and NumPy** processing with quality validation, improving usable data yield by **36%** across **4 dataset categories**.
- Ran **rapid experiments** testing model hypotheses, iterating quickly on promising results and chasing incremental model improvements, reducing experiment iteration cycle time by **34%** across **11 model variants**.
- Turned successful experiments into **reusable Python infrastructure and pipelines**, automated **CI/CD via GitHub Actions**, reducing release cycle from **24 min to 8 min** across **11 research-to-production pipelines**.

Software Engineer Intern
Lumina View

May 2025 – Aug 2025
New York, NY

- Built **Python PyTorch time-series forecasting models** over large, imperfect ground-truth datasets, debugging nontrivial ML systems, improving forecast accuracy by **50%** across **95K+** monthly prediction cycles.
- Built **dependable data pipeline abstractions** for combining multi-source geospatial and time-series datasets with inconsistent quality and coverage, reducing data cleaning effort by **41%** across **4 data domains**.
- Crossed boundaries between research and engineering, forming hypotheses and designing experiments, maintaining **99.3%** model reliability while iterating on **9 research directions**.

Software Engineer I
Mavdero Techservices

Dec 2023 – Aug 2024
Chennai, India

- Built **Python PyTorch and Scikit-learn models** over large structured and unstructured datasets, processing **95K+** daily records across **5 production model pipelines**, maintaining p95 inference latency under **140ms**.
- Cleaned and combined **messy datasets with inconsistent formats**, achieving **91.4%** data quality accuracy, reducing false-positive anomaly detection rate by **28%** through systematic validation.
- Turned research experiments into **reusable production infrastructure**, automated **CI/CD via Jenkins**, reduced deployment failures by **32%** across containerized **Docker** environments.

Projects

Multi-Source Weather Data Assimilation and Forecasting Platform
(SageMaker, S3, EMR), PostgreSQL, Docker

Python, PyTorch, Xarray, NetCDF, AWS

- Built a **Python PyTorch data assimilation platform** incorporating observations from **multiple sensor and satellite-style sources**, cleaning and combining **500K+** simulated daily geospatial and time-series records with inconsistent formats, improving usable data quality by **36%** across **3 data source types**.
- Ran rapid experiments testing model architectures, turned successful results into **reusable Python infrastructure**, enforced **GitHub Actions CI/CD** quality gates, achieving **99.2%** pipeline reliability across all evaluation runs.

Time-Series Forecasting and Foundation Model Research System *Python, PyTorch, TensorFlow, Pandas, NumPy, AWS (SageMaker, S3, CloudWatch), GitHub Actions*

- Designed a **Python PyTorch foundation model** for time-series forecasting across multiple weather-related datasets, processing **95K+** daily prediction cycles with a **99.3%** reliability rate through systematic experiment design.
- Built **reusable experiment infrastructure** accelerating future research iteration, enforced **automated testing** via **GitHub Actions** that reduced post-deploy regressions by **27%**, sustaining **30 days** of stable model operation.