Ishaan Potle

+1 631-615-5837 | ishaansantosh.potle@stonybrook.edu | Linkedin | Github

Education

Stony Brook University — Stony Brook, New York
Master of Science in Data Science — *Expected Graduation 2026*
Coursework:Data Analysis, Data Management, Statistical Learning, Statistical Computing, Visualization

NMIMS's, MPSTME — Mumbai, India
Bachelor of Technology in Computer Engineering — *2024*

Technical Skills

Languages: Python, R, SQL
ML/AI: PyTorch, TensorFlow, Transformers, SHAP, XGBoost, Scikit-learn, Fairlearn
Infra: Docker, Redis, FastAPI, Celery, GitHub Actions, Streamlit, Qdrant
Concepts: GNNs, LLMs, RAG, MLOps, Explainable AI, Federated Learning, NLP

Experience

Technical Program Management Intern — May. 2025 – Aug. 2025
Quantum Pulse Consulting — *Remote, United States*

- Drafted 4 technical specifications templates and PRDs to define features for LLM and analytics projects.
- Collaborated with 8 engineers and 2 PMs to align project OKRs, leading to a GenAI prototype demo completion.
- Worked with CTO to prioritize roadmap features based on model performance,impact and technical feasibility.
- Authored internal documentation (15+ pages) clarifying workflows, increasing onboarding for new interns by 20%.

Data Science Intern — May. 2025 – Aug. 2025
Quantum Pulse Consulting — *Remote, United States*

- Improved model testing workflows by building 6 API endpoints using FastAPI, cutting testing time by 25%.
- Processed and cleaned over 200,000 rows of time-series data contributing to model training pipelines.
- Benchmarked 3 pre-trained models (Hugging Face + scikit-learn) on sample datasets, helping select the optimal baseline for internal NLP tasks.

Projects

Real-Time AML Fraud Detection Engine | *PyTorch Geometric, DGL, FastAPI, Redis* 2025

- Developed an AML detection engine using Graph Neural Networks (GraphSAGE, GAT, TGN) for real-time analysis of cryptocurrency transaction networks. Achieved 90% faster inference (100–500ms vs 2–5s).
- Built a microservices architecture with FastAPI backend, Streamlit dashboard, and Dockerized deployment; integrated Prometheus, Grafana, and ELK for observability with 71+ system metrics.
- Engineered temporal graph processing for time-series fraud analysis; added explainable AI modules using SHAP and Captum for regulatory interpretability

Bias-Resistant Resume Ranker | *Redis, Sentence Transformers, Llama 3, Fairlearn* 2025

- Constructed an AI-driven platform to eliminate bias in resume screening and deliver fair candidate evaluation and resume improvement tools with resume quality scoring system detailed explanations for each score.
- Implemented semantic skill matching using Sentence Transformers and integrated a local Llama 3 model for advanced technical and soft skill extraction beyond keyword matching.
- Enabled batch DEI fairness auditing using demographic parity and equalized odds metrics to analyze bias at scale.

Federated Learning System for Health Risk Prediction | *PyTorch, Flower, Azure, Docker* 2025

- Designed a privacy-preserving federated learning system enabling distributed model training across multiple clients for health risk prediction using real-world health datasets from FHIR and Fitabase.
- Made federated averaging and client-server communication protocols using the Flower framework.
- Automated feature importance with SHAP, producing per-client and aggregate explainability metrics.

Mistral-Med: Local LLM Medical Assistant with Vector Search | *Mistral-7B, Qdrant, RAG, SHAP* 2025

- Created a domain-specific medical chatbot leveraging Mistral-7B and RAG with vector search(Qdrant), contextual document retrieval(PubMed, MedQuAD), and LLM generation for explainable medical question answering.
- Deployed quantized LLMs (GGUF) on local and cloud infrastructure, optimizing for inference speed, memory usage, and responsiveness.

Manuscript

- "Ensemble Learning with Hybrid Modeling for Multivariate AQI PM2.5 and PM10 Forecasting in Mumbai", International Conference on Computer Science and Emerging Technologies, IEEE
- "Enhancing Disaster Tweet Classification with Ensemble Models and Multiple Embeddings",2023 4th Global Conference for Advancement in Technology (GCAT), IEEE