LAVYA MIDHA

+1 (413) 800-1278 | lavyamidha015@gmail.com | Linkedin | Portfolio | Github

EDUCATION

New York University New York, NY
Master of Science in Data Science (Industry Concentration) Expected May 2026
Cumulative GPA: 3.762/4.0
Relevant Coursework: Probability, Statistics, Text as Data, Big Data, Machine Learning, Practical Training for DS, ML in Finance.

Boston University Boston, MA
Bachelor of Science in Data Science, Minor: Statistics Sepl 2021 - May 2024
Cumulative GPA: 3.75/4.0
Relevant Coursework: Data Visualization, Machine Learning and AI, Algorithms, Statistical Modeling, Data Mechanics, Optimization, NLP, Regression Analysis, Inference, Analysis of Variance.

WORK EXPERIENCE

ReferU.AI Lewes, Delaware
Machine Learning Engineer Intern May 2025 – Present

- End-to-end GenAI Legal Assistant: Designed and deployed a production-grade, multi-agent RAG system (LangChain + OpenAI + GCP) for citation-aware legal retrieval, integrating retrieval → reranking → guardrail pipelines with Dockerized FastAPI microservices and CI/CD automation.
- Retrieval Scoring Framework: Built a scalable benchmarking and evaluation pipeline for court opinion evaluation (GCP + CourtListener API) to measure precision/recall, latency, and coverage; driving data-driven optimization across 10K+ legal citations.
- Adversarial Validation Layer: Engineered a parallel multi-agent QA framework (Validator, Auditor, Bias Hunter, Edge-Case Challenger) to reduce hallucinations and bias, boosting factual accuracy +22% and cutting manual review effort 35%

BU Spark! Boston, MA
Project Manager Intern Jan 2024 – May 2024

- Directed 3 applied ML teams delivering forecasting and optimization tools for civic-tech clients; acted as the primary client liaison, translating non-technical policy goals into ML problem statements and deployment plans.
- Collaborated with teams to build Tableau dashboards that converted complex ML outputs into intuitive, non-technical insights for city officials and executives, enabling data-driven decision-making.
- Improved iteration speed by 14% by introducing agile pipelines and structured stakeholder review loops, ensuring deployed models were interpretable, accurate, and aligned with end-user needs.

Boston University School of Public Health Boston, MA
Data Science Research Intern Feb 2023 - Dec 2023

- Informed $MM+ infrastructure investments by building time-series models on air quality data from 125+ schools, prioritizing equity-first capital allocation.
- Surfaced systemic disparities in classroom environments by designing an OpenCV pipeline to extract spatial features (window access, density) from floor plans.
- Cut anomaly response times by 35% by deploying a real-time dashboard flagging temperature spikes for administrators and facilities teams.

SELECTED PROJECTS

- **Probability of Default Prediction System:** Created an end-to-end PD modeling system for SME credit risk, integrating extensive ratio engineering, XGBoost modeling, and advanced calibration (Laplace, bucket mapping) to produce accurate, well-behaved PDs validated via OOT splits and AUC/KS/Brier diagnostics.
- **Voice PrepAI:** Developed an LLM-driven interview preparation agent with real-time audio streaming and multi-agent evaluation that provides personalized, resume-, role-, and company-aware feedback to improve answer quality and structure.
- **Bitcoin Accumulation (for Trilemma Foundation):** Built a multi-agent system for bitcoin accumulation in which learning-based agents detect market regimes, forecast short-term risk, and adjust DCA schedules in real time to maximize accumulated BTC across varying market conditions.

SKILLS

Programming & Data: Python (Pandas, NumPy, scikit-learn, matplotlib) , SQL, C++, Java, R, TypeScript, Node.js.
ML/ AI: PyTorch, TensorFlow, LightGBM, XGBoost, Transformers, Recommender Systems, Time Series, NLP, Computer Vision, GenAI & Multi-Agent Systems, LangChain.
Data Infrastructure: Spark, Hadoop ,Dataproc, Docker ,REST APIs, ETL, HPC, JIRA.
Cloud & Deployment: GCP, AWS, Azure, Google Ads, Google Analytics ; CI/CD, Monitoring.
Visualization & Communication: Tableau, PowerBI, Looker Studio Python Panel, Rshiny Stakeholder-Facing, Agile.