

Lavya Midha

✉ lavyamidha015@gmail.com | 📞 413-800-1278

🐙 [GitHub](#) | [LinkedIn](#) | 🌐 [Portfolio](#)

Education

New York University, New York, NY

M.S. in Data Science (Industry Concentration)

Expected May 2026

GPA: 3.762/4

Relevant Coursework: Probability, Statistics, Text as Data, Big Data, Machine Learning, Practical Training for DS, Applied ML in Finance.

Leadership: Events – Graduate Student Community Building Group at NYU CDS

Boston University, Boston, MA

September 2021 - May 2024

B.S. in Data Science, Minor: Statistics

GPA: 3.75/4

Relevant Coursework: Data Visualization, Machine Learning and AI, Algorithms, Statistical Modeling, Probability, Big Data, Data Mechanics, Optimization, Regression Analysis, NLP, Inference, Analysis of Variance.

Leadership and Awards: Undergraduate Research Grant, Dean's List Awardee, Secretary – BU Data Science Association, International Peer Mentor, Peer Tutor, Admissions Ambassador

Recognition: Featured in BU Spark! Demo Day for civic tech innovation, recognized for public impact work in data science ([Link](#))

Work Experience

ReferU.AI, Lewes, Delaware

May 2025 – Present

Machine Learning Engineer Intern

- **GenAI Legal Assistant (95%+ accuracy):** Built and deployed an **end-to-end multi-agent system** (LangChain + OpenAI + vector DBs) for citation-aware case retrieval. Designed retrieval → reranking → guardrails pipeline, and developed an **in-house retrieval scoring system (GCP + CourtListener)** to benchmark precision/recall at scale.
- **Productionized GenAI system:** Exposed APIs via **FastAPI + Docker on GCP** with CI/CD, monitoring, and dashboards; orchestrated parsing, summarization, and citation-validation agents to **reduce hallucinations and scale reliability** in live environments.
- **Ad Intelligence Stack (+15% ROI):** Partnered with marketing executives to deliver a **production ML pipeline** (LightGBM + embeddings) for real-time keyword ranking, boosting nationwide Google Ads ROI by **15%+**. Built **jurisdiction-aware dashboards** (Looker Studio) integrating court dockets, demographics, and Ads KPIs to drive campaign optimization.
- Python, GCP, LightGBM, OpenAI APIs, LangChain, Looker Studio, Google Ads, Google Analytics, GenAI, Agents, Vector DBs.

BU Spark!, Boston, Massachusetts

January 2024 - May 2024

Project Manager

- **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.
- Tableau, MLOps, Python, Google BigQuery, Agile, Stakeholder Engagement.

BU School of Public Health, Boston, Massachusetts

February 2023 - December 2023

Data Science Research Assistant

- **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 that flagged temperature spikes for administrators and facilities teams.
- Python, SQL, R, Python Panel, scikit-learn, OpenCV, Infrastructure Analytics, R Shiny.

Selected Projects

- **VendorRank (PyramidAI):** Built a FastAPI-based analytics engine ranking **1,000+ CRM** tools using KeyBERT, SBERT, and weighted similarity scoring. **Delivered high-relevance vendor recommendations** that improved enterprise procurement decisions.
- **MovieLens Recommender (33M ratings):** Designed a collaborative filtering + MinHash + ALS pipeline with tag-genome regression, validated with ranking metrics (Precision@100, MAP, NDCG). Demonstrated scalable personalization **for investor clustering and advisory use cases**.
- **Capital Misallocation Detection (RiseUp):** Identified **\$40M+ inefficiencies** in Boston Police budgets via anomaly detection and feature engineering. Produced **quantified corrective strategies** for civic leaders, informing financial allocation at city scale.

Skills

- **Programming & Data:** Python (Pandas, NumPy, scikit-learn, matplotlib), SQL, C++, Java, R, TypeScript, Node.js
- **ML/ AI:** PyTorch, TensorFlow, LightGBM, XGBoost, Transformers (BERT, RoBERTa, BART), Recommender Systems (ALS, MinHash), Time Series, NLP, Computer Vision, GenAI & Multi-Agent Systems
- **Data Infrastructure:** Spark, Hadoop, Dataproc, Docker, REST APIs, ETL, HPC
- **Cloud & Deployment:** GCP (BigQuery, Looker Studio, Cloud Run/Build), AWS, Azure; CI/CD, Monitoring
- **Visualization & Communication:** Tableau, Power BI, Python Panel, Stakeholder-Facing Dashboards