Daivya Shah

New York, NY | (917) 847-4174 | daivya.shah@nyu.edu | linkedin.com/in/daivya-shah/ | github.com/Daivya-Shah

EDUCATION

New York University, Center for Data Science *Expected Graduation: May 2026*

B.A. in Computer Science & Data Science, Minor in Business

- **Relevant Coursework:** Fundamentals of Machine Learning, Data Management/Analysis, Principles of Data Science, Causal Inference, Regression and Forecasting Models, Probability and Statistics, Linear Algebra, Data Structures, Calculus III

PROFESSIONAL EXPERIENCE

Newmark Group **New York, NY**

Software Engineer Intern | React, FastAPI, Databricks, SQL Server, React Query, IndexedDB *Jun 2025 – Aug 2025*

- Built a full-stack Goal and Resource Management Dashboard that powered hierarchical OKRs, employee allocations, and CapEx/OpEx forecasting, adopted by stakeholders, leveraging **React/FastAPI**, SQL APIs, and **IndexedDB** caching.
- Developed a responsive CRE Stacking Plan in **React/TypeScript** with a CSV pipeline powered by **Databricks** to transform **10K+** leasing records into structured tenant KPIs and floor-plan visualizations, reducing reporting time significantly.
- Engineered a Deal Pipeline Kanban with @dnd-kit and **React Query**, normalizing state models to track **$500M+** in active deals and deliver portfolio value and stage velocity insights for leadership to monitor deal flow.
- Built a GIS Dashboard with Zod-validated request forms and automated email workflows for geospatial report requests.

Everise **New York, NY**

Data Analyst Intern | Apache Airflow, SQL Server, Pandas, NumPy, Statsmodels, PowerBI *Jul 2024 – Aug 2024*

- Automated **ETL** workflows with **Apache Airflow**, integrating **500K+ data points** from **6+** data sources into a centralized data warehouse (MISDW), and delivered **2 real-time Power BI dashboards** adopted by stakeholders for decision-making.
- Conducted **root cause analysis** on customer service KPIs across **100K+** interactions using SQL queries, Python, and **DAX**; identified peak-hour call surges and escalation delays, implementing fixes that raised Customer Satisfaction scores by **7%**.
- Built **ARIMA-based** predictive models in Python achieving **92%** forecast accuracy on billing data for enterprise clients, enabling finance teams to anticipate fluctuations in a **$20M+** annual Bill-to-Pay cycle and mitigate billing risks.

eMeasurematics - Industrial Autonomous Solutions **Chicago, IL**

Machine Learning Intern | Flask, SQLAlchemy, scikit-learn, Random Forests, MySQL, Tableau *Jun 2024 – Jul 2024*

- Developed a **Flask** web app for predictive maintenance of Kress industrial vehicles, integrating historical and user-entered data via **SQLAlchemy** and deploying a **Random Forest** model that predicted failures with **85%** accuracy.
- Built a **Tableau** dashboard with **MySQL** for predictive maintenance, analyzing **50K+** sensor logs to optimize scheduling.
- Configured, calibrated, and updated firmware on 16 radar sensors for precise slab positioning across 4 steel facility cranes.

LEADERSHIP & PROJECTS *(See more at daivyashah.com)*

NYU Machine Learning Club | *Vice President* *Jun 2024 – Present*

- Expanded membership to **800+** and increased active participation to **145+** by organizing events and community programs.
- Organized 8+ industry expert talks and hiring events with speakers from Meta, Datadog, Salesforce, Pinterest, and David Lin (ex-CTO of JPMorgan, CEO of Linvest21), providing career insights to an audience of 100+ attendees.

Fitzty | *Fashion Intelligence Platform* *Jun 2025 – Aug 2025*

- Built a **computer vision** system using GPT-4 Vision and DALL-E 3 to analyze clothing images, extract visual features, and generate outfit recommendations stored in a **PostgreSQL-Supabase** database.
- Developed a data-driven **recommendation model** that learned from user activity to improve outfit relevance and automate fashion analysis across **10K+** wardrobe items.

Harvest | *AI-Powered Manufacturer Discovery Platform* *Jan 2025 – Mar 2025*

- Led development of a **Retrieval-Augmented Generation (RAG)** system that processed data from **500+** manufacturer catalogs using **Selenium**, NLP, and OCR, organizing **100K+** records in **MongoDB.**
- Designed the retrieval engine with OpenAI embeddings, FAISS, and GPT-4 to match suppliers by product type and capacity, achieving **90%** accuracy and cutting discovery time from weeks to days.

Trustworthy AI Lab UCLA | *GenAI Hackathon* *Jun 2024 – Jul 2024*

- **Won 1st place out of 30+ university teams** by building a Data Clean Room (DCR) with Azure Confidential VMs, Key Vault, and TPM2-tools, creating a secure environment for multi-party data sharing and predictive analytics.
- Increased CTR prediction accuracy by **20%**, outperforming the baseline model, by integrating heterogeneous datasets and synthetic data into Random Forest models while maintaining full confidentiality.
- Designed and trained **GANs** to generate synthetic datasets with **99.87% fidelity**, enhancing overall model performance.

TECHNICAL SKILLS

Skills/Frameworks: Python, Java, C++, R, SQL, JavaScript, TypeScript, Node.js, Next.js, React, Flask, PyTorch, TensorFlow, Scikit-Learn, Pandas, GANs, MySQL, MongoDB, SQLAlchemy, AWS, Docker, Tableau, Power BI, OpenAI, Pinecone, Git/GitHub