

# Indeera Saha

indeera.saha@nyu.edu | +1 (678)-683-5308 | New York, New York  
[indeerasaha.github.io](https://github.com/indeerasaha) | [linkedin.com/in/indeerasaha](https://linkedin.com/in/indeerasaha) | [github.com/indeerasaha](https://github.com/indeerasaha)

## EDUCATION

### New York University

*B.S. in Computer Science, Minor in Business Studies*

Courses: Machine Learning, Algorithms, Computer Networking, Databases, Container Operating Systems, Regression & Forecasting, Algorithmic Problem Solving, Software Engineering, Management, Financial Accounting, Corporate Finance

Manhattan, NY

Sep. 2022 – May 2026

## EXPERIENCE

### Data Science Intern

*Acadia Healthcare*

June 2024 – Present

*Remote*

- Built long-context LLM summarization API (200K+ tokens) with tournament-pivot Top-K ranking, achieving 10x speedup over original heap-based approach
- Automated synthetic data pipeline generating 10M+ tokens per batch for fine-tuning domain-specific LLMs using Mistral-7B, GPT-4.1, and Snowflake Cortex
- Developed internal Python package index with ML utilities and Snowflake integrations (logging, CI/CD, access control), adopted across Data & Analytics team

### Research Assistant

*NYU Urban Modeling Group w/ Prof. Debra Laefer*

June 2023 – Present

*Brooklyn, NY*

- Developed a predictive algorithm to reconstruct underground pipe networks from hydrant coordinates (92% F1-score)
- Applied Hough Transform and Random Forest/SVM Computer Vision models to extract 2.3K+ hydrant points from a 2km<sup>2</sup> Lower Manhattan raster image, achieving 73% detection accuracy (1.5x above industry baseline)
- Optimized networks with Delaunay Triangulation for connectivity and a Genetic Algorithm for routing

### Club President

*Tech@NYU*

Jan. 2023 – May 2025

*Manhattan, NY*

- Managed a 400+ member student club, across event logistics, communications, 4 flagship programs, finances, and collaborations with Anthropic, Notion, Adobe, Vercel, Databricks, Brave Browser, etc.
- Implemented data-driven strategies, doubling events (from 8 to 16+ per sem) and boosting membership by 32%
- Launched an open-source program training students to build full-stack applications, attracting industry interest
- Organized and moderated a 2023 'Women in Tech Panel' with founders from Resy, NYU start-ups, and AI+Fashion

## PROJECTS

### Script Continuity RAG System | *Langchain, Pinecone, Pandas, Giskard, SQLite*

Aug. 2025 – Present

- Built a RAG system for TV script continuity detection with custom scene-based chunking and metadata retrieval, improving LLM rewriter performance by 13.7% over standard splitting
- Engineered PDF parsing pipeline with regex extraction and SQLite storage for structured queries across 200+ scripts
- Implemented chain-of-thought prompting with few-shot examples & Giskard evaluation framework for Q&A validation

### Feline Traits & Care Analysis | *Selenium, BeautifulSoup, scikit-learn, Pandas, SciPy*

Aug. 2024 – Present

- Scraped and cleaned 4,000+ cat adoption records, performing missing value imputation & categorical standardization
- Conducted exploratory data analysis and statistical correlation testing, identifying moderate positive correlation between cat age and grooming needs
- Developed Random Forest classification model to predict grooming needs from physical and behavioral traits, achieving 87% accuracy and identifying coat length as primary predictor through feature importance analysis

## TECHNICAL SKILLS

**Languages:** Python, C++/C, SQL

**Developer Tools:** Git, Snowflake, GCP, Docker, Kubernetes, Microsoft Azure, Databricks, VS Code, Linux / Bash

**Libraries:** Pandas, NumPy, OpenCV, Matplotlib, LangChain, Pinecone, Shapely, Flask, SQLAlchemy, scikit-learn