SNEHA DHARNE

+1(201)204-2708 | New York Metropolitan Area | snehadattadharne@gmail.com | [LinkedIn](https://www.linkedin.com/in/sneha-dharne-6837441b6/) | [GitHub](https://github.com/SnehaDharne) | [Portfolio](https://snehadharne.vercel.app/)

# EDUCATION

## Stevens Institute of Technology, MS Computer Science September 2023 – May 2025

## Courses: Machine Learning, Deep Learning, NLP, Big Data Technologies, Business Applications of Gen AI 4.0/4.0

## Manipal Institute of Technology, BTech Information Technology, Minor in Big Data Analytics July 2019 – July 2023

## Courses: Database Systems, Data Warehousing and Data Mining, Big Data Integration and Processing 3.62/4.0

## SKILLS

# Languages and Frameworks: C#, C++, Java, Python, JavaScript, PySpark, MlLib, NodeJS, ReactJS, Git, Postman, FastAPI

**Gen AI and cloud tools**: LangChain, LLMs, Prompt Engineering, OpenAI, Google Cloud, AWS

**Data Science**: SQL, Power BI, Pandas, NumPy, TensorFlow, Matplotlib, Seaborn, Excel, MS Office Suite

## EXPERIENCE

## Oncology Reference Inc – *Data Engineering and DevOps Intern | Jersey City, USA* September 2024 – Present

* Built ETL pipelines for unstructured data, automating **categorization of documents using NLP,** reducing processing time by 30%
* Architected and developed **data pipelines** from 14 other medical data sources, automating **cleaning, preprocessing** and **big data updates in MongoDB** and reducing manual intervention by **60%**
* Developed a data labeling system using **NER and GPT4o,** ensuring accurate data tagging for clinical analysis
* Developed “matching” **mini games** for medical writers for manual labeling, cutting labeling time by **40%** and boosting efficiency
* Designed **dashboards** for research and drugs statistics, making **data analysis 50% faster** and simplifying insights for researchers
* Implemented **SQL +** **RAG-based pipelines** for efficient data retrieval, validated by researchers for smooth and reliable usage
* Assisted in **optimizing website performance on GCP**, improving load times and user accessibility
* Led scrums on **defining data objectives**, translating user stories to technical requirements, and ensuring projects met deadlines

## Stevens Institute of Technology (OneIT) – *Technical Consultant | Hoboken, USA* March 2024 – September 2024

* Provided technical support to 700+ university users, resolving complex data and security issues efficiently.
* Communicated **technical solutions clearly to non-technical us**ers, improving user satisfaction and experience.
* Assisted campus offices, staff, and faculty with **hardware/software installations, maintenance, and troubleshooting**.
* Recognized by OneIT leadership, awarded **MVP and Speed Racer for exceptional performance and efficiency**

## Deloitte USI – *Business Analyst Intern | Hyderabad, India* May 2022 – July 2022

* Implemented **DAX functions to** transform Nike sales data, improving data processing efficiency by 50%
* Optimized data models in **Power BI** by normalizing data into **3NF,** defined relationships and computed **Key Performance Indicators** (KPIs) - Year-to-Date (YTD) category performance, growth opportunities and market share. Revealed a 40% revenue increase despite a 60% drop in unit sales annually from 2012 - 2014
* Ensured secure and scalable user data and transaction log management with **SQL** for a web app

## PROJECTS

## [Microsoft Azure (PyRit)](https://github.com/Azure/PyRIT) – Open-Source Contribution

Integrated harmful prompt datasets into Hugging Face’s LLM to improve model robustness and documentation clarity

## [Chubb (Capstone Project) – Real-time Big Data Pipeline for Stock Analytics](https://github.com/SnehaDharne/StockAnalyticswithAWS)

## Developed a real-time stock data pipeline using AWS, Spark, and Kafka, optimizing data processing, tracking metadata, and enabling real-time analytics with Streamlit visualizations. Reduced processing time by 30%, improving decision-making speed

## [Fin-AI Co-pilot (Financial Report generator)](https://github.com/SnehaDharne/FinancialReportGeneratorGenAI)

## Built an AI tool using LangChain to automate financial data extraction with chain-of-thought reasoning, enabling interactive dashboards and report generation. Integrated agents like calculators, graph and CSV generators, reducing report creation time from 8 hours to 10 minutes

## [Symptom Extraction with Gemini Pro (VAERS Dataset)](https://github.com/SnehaDharne/TemporalAndAssociativeRelationships-VAERS)

## Built a big data-driven solution leveraging LLMs to extract symptoms from unstructured VAERS data, optimizing batch processing and implementing advanced prompt engineering. Enabled temporal and associative analytics, providing actionable insights for researchers to better understand symptom patterns.

## [ETL and Predictive Modeling (nyc.gov data)](https://github.com/SnehaDharne/nyc.gov-data-analytics)

## Led ETL on 5M records across 3 datasets using PySpark, performing feature engineering and integrating spatial data (latitudes/longitudes) to predict collision risk. Utilized clustering to identify hotspots, built a prediction model with 84% accuracy, and scaled the solution on GCP, reducing training time by 75%

## LEADERSHIP

## Stevens Institute of Technology (GSL) – *Graduate Peer Leader | Hoboken, USA* September 2024 – December 2024

Implemented strategies to assist international student peers with challenges of studying abroad, promoting cultural exchange and academic success.