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

## GPA: 4.0/4.0

## Capstone Project: Chubb. Big Data Analytics on financial data.

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

## GPA: 3.62/4.0

## Project: CNN based image-classification tool to detect early-stage Chronic Ocular Diseases

## SKILLS

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

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

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

## EXPERIENCE

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

* Built automated **ETL pipelines for unstructured data**, categorizing research documents using **NLP**
* **Optimized MongoDB workflows**, automating large dataset updates, resulting in a **20% reduction** in query times
* Developed and optimized Cloud SQL databases for complex relationship queries

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

* Delivered expert technical consultation to over 700+ university users, resolving complex data and security issues
* Troubleshot and **effectively communicated technical solutions to non-technical users**, enhancing user satisfaction
* Recognized by the Director and Associate Director of OneIT, awarded **MVP and Speed Racer employee in OneIT**

## Deloitte USI – *Business Analyst Intern* Hyderabad, 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

## [Fin-AI Co-pilot (Financial Report generator)](https://github.com/SnehaDharne/FinancialReportGeneratorGenAI) September 2024 – December 2024

* **Built an AI-driven financial analysis tool** leveraging LangChain to extract, analyze, and visualize data from 10-K financial reports
* **Designed and deployed modular workflows** including custom tools, AI agents, and data pipelines to process and transform financial data into actionable metrics and interactive graphs
* **Streamlined data extraction by 40%** through automation of SEC 10-K processing into JSON formats, integrating NLP-driven agents for enhanced user query responses

## [Symptom Extraction with Gemini Pro (VAERS Dataset)](https://github.com/SnehaDharne/TemporalAndAssociativeRelationships-VAERS) September 2024 – December 2024

* Developed a big data, LLM-driven solution to extract insights from unstructured free-text data in the VAERS dataset.
* Optimized batch processing and prompt engineering, enabling temporal and associative analytics on symptoms to derive actionable insights

## [Microsoft Azure (PyRit)](https://github.com/Azure/PyRIT) October 2024

## Integrated harmful prompts datasets from Lateral Adversarial Training LLM model from hugging face

## Tested dependencies and availability of data within the model and improved project documentation

## Chubb (Capstone Project) – Real-time Data Pipeline for Stock Analytics March 2024 – May 2024

## Developed a real-time stock data pipeline using AWS EMR and S3, Kafka, and PySpark for processing tickers’ data, implemented micro batch processing and parallel processing for efficient streaming data ingestion

## Set up Spark Structured Streaming to compute KPIs, and created visualizations with Streamlit

## Harvested event logs and live metadata stream, maintaining a Data Quality Index (DQI) of 0.9

## [ETL and Predictive Modeling (nyc.gov data)](https://github.com/SnehaDharne/nyc.gov-data-analytics) March 2024 – May 2024

* Led **ETL** processes on **5 million records** from MV Collision data, integrating data from multiple sources
* Leveraged **Spatial Clustering** on location coordinates to detect hotspots and get a **severity score** between 0 to 1
* Incorporated spatial data and data-driven insights to build a collision risk prediction model achieving 84% accuracy
* Scaled out executed on **Google Cloud Platform** resulting in 75% drop in training time

## LEADERSHIP

* **GRADUATE PEER LEADER,** Stevens Institute of Technology - Implemented strategies to assist international student peers with challenges of studying abroad, promoting cultural exchange and academic success.
* Volunteered for door-to-door fundraising to support medical expenses at **THE NATIONAL ASSOCIATION FOR SPECIALLY ABLED ENTERPRISES, MUMBAI**