SNEHA DHARNE

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

Highly motivated MS Computer Science candidate seeking an internship or entry-level role at in Data or ML. Proven ability to develop and implement data engineering and machine learning solutions using Python, PySpark, and cloud technologies.

# 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: Python, C++, Java, JavaScript, PySpark, Git, Postman

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

**Gen AI and cloud tools**: Prompt Engineering, OpenAI, AWS (EC2, S3, Lambda)

## 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

* Troubleshot and resolved an average of 20 data security and technical support issues per week, aiding 700+ university users
* 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

## PROJECTS

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

* Processed 5M+ data points using PySpark ETL workflows, including complex join operations across three datasets.
* 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
* Scale out executed on **Google Cloud Platform’s DataProc** resulting in 75% drop in training time

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

* Built an AI-powered financial report generator using **LangChain and OpenAI**, automating the extraction of financial data from unstructured sources and **enabling chain-of-thought prompting and reasoning** for accurate insights
* Integrated interactive dashboards with tools **like CSV generators and graph plotters**, allowing users to visualize key metrics such as revenue, expenses, and profit trends in real-time
* Automated financial report generation, reducing processing time from 8 hours to 10 minutes, increasing analyst productivity

[**Chubb (Capstone Project) – Real time data analytics with AWS**](https://github.com/SnehaDharne/StockAnalyticswithAWS)

* Built a real-time stock data pipeline with **AWS (S3, EC2), PySpark, and Kafka**, reducing data lag by 80% for faster insights
* Implemented **Kafka producer-consumer architecture** to for streaming data from yfinance, computed **KPIs with Spark Structured Streaming** and **parallel processing**, incorporated **metadata logging** for proactive monitoring
* Prototyped a **predictive model using minute-level price differences** to forecast short-term stock movements

## PUBLICATIONS

**A machine learning-based clinical decision support system for diagnosing GDM.** DOI: [10.1016/j.jaim.2024.101051](https://www.sciencedirect.com/science/article/pii/S0975947624001669?via%3Dihub)

Contributions: Developed and implemented the machine learning algorithms for the clinical decision support system. Achieved an F1 score of 0.84, optimized feature selection to identify top 3 features, and evaluated model reliability through accuracy and correlation analysis.