# SNEHA DHARNE

+1(201) 204-2708 | New York Metropolitan Area | snehadattadharne@gmail.com | LinkedIn | GitHub | Portfolio

## EDUCATION

**Stevens Institute of Technology, MS Computer Science** September 2023 – May 2025
**Relevant Coursework:** Machine Learning, Deep Learning, NLP, Big Data Technologies **4.0/4.0**
**Manipal Institute of Technology, BTech Information Technology, Minor in Big Data Analytics** July 2019 – July 2023
**Relevant Course:** Database Systems, Data Warehousing and Data Mining, Big Data Integration **3.62/4.0**

## SKILLS

**Data Analysis & Visualization**: SQL, Power BI, Tableau, Excel, Pandas, NumPy, Matplotlib, Seaborn
**Programming & Tools**: Python, R, Java, JavaScript, Git, Postman
**Machine Learning & Big Data**: Scikit-learn, TensorFlow, PyTorch, Spark, Kafka, PySpark, AWS (S3, EC2, Lambda)

## EXPERIENCE

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

- Automated unstructured data categorization with NLP-driven ETL pipelines, reducing processing time by 30%.
- Architected and developed **data pipelines** from **14 other medical data sources,** automating **cleaning, preprocessing** and **data updates in MongoDB** and reducing manual intervention by **60%**
- Designed **dashboards** for clinical developments, 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
- Developed a **data labeling system** using **NER and GPT4o,** ensuring accurate data tagging for clinical analysis

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

- Resolved an average of 20 data security and technical support issues per week, aiding 700+ university users
- Communicated **technical solutions clearly to non-technical users**, improving user satisfaction and experience
- Assisted campus offices, staff, and faculty with **hardware/software installations, maintenance, and troubleshooting**
- Recognized by OneIT leadership, "**MVP" and "Speed Racer Employee" for two months**

**Deloitte USI** – *Business Intelligence 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 **KPIs**
- Identified a 40% revenue increase (2012-2014) despite a 60% drop in unit sales through in-depth data analysis
- Ensured secure and scalable user data and transaction log management with **SQL** for a web app

## PROJECTS

**ETL and Predictive Modeling (nyc.gov data)**

- 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)**

- 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**

- 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

Shetty, N.P., Shetty, J., Hegde, V., Dharne, S.D., & Kv, M. (2024). A machine learning-based clinical decision support system for effective stratification of gestational diabetes mellitus and management through Ayurveda. *Journal of Ayurveda and Integrative Medicine*, 15(6), 101051. PMC: PubMed) DOI: 10.1016/j.jaim.2024.101051

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.