Sneha Dharne

New York, NY

201-204-2708 snehadattadharne@gmail.com linkedin.com/in/snehadharne github.com/snehadharne

Education

Stevens Institute of Technology, MS in Computer Science **Sep. 2023 – May 2025**
Relevant Coursework: Machine Learning, Deep Learning, NLP, Big Data Technologies *4.0/4.0*

Manipal Institute of Technology, BTech in Information Technology **Jul. 2019 – Jul. 2023**
Relevant Coursework: Database Systems, Data Warehousing and Data Mining, Big Data Integration *3.62/4.0*

Experience

Oncology Reference Inc. **Sep. 2024 – Present**
Data Engineering and Analytics Intern *Jersey City, NJ*

- Automated unstructured data categorization with NLP-driven ETL pipelines, reducing processing time by 30%.
- Architected and developed data pipelines from 14 medical data sources, automating cleaning, preprocessing and data updates in MongoDB, 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 **Mar. 2024 – Sep. 2024**
Technology Consultant *Hoboken, NJ*

- 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 as 'MVP' and 'Speed Racer Employee' for two months

Deloitte USI **May. 2022 – Jul. 2022**
Business Intelligence Analyst Intern *Remote, India*

- 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
- Recognized by Deloitte USI Leadership, converted to full time offer for Business Analyst Position

Projects

Fin-AI Co-pilot (Financial Report generator) | *LangChain, OpenAI, StreamLit* **Sep. 2024 – Dec. 2024**

- 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

ETL and Predictive Modeling (nyc.gov data) | *PySpark, Machine Learning, Google Cloud Console* **Mar. 2024 – May 2024**

- 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 Plaform's DataProc resulting in 75% drop in training time

Chubb (Capstone Project) – Real time data analytics with AWS | *Python, PySpark, Kafka, AWS* **Mar. 2024 – May 2024**

- 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

Technical Skills

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

Publications

Machine Learning Research with Healthcare Data **Nov 2022 – Dec 2022**
Journal of Ayurveda and Integrative Medicine (PubMed Indexed) *Manipal Institute of Technology*

- 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. DOI: 10.1016/j.jaim.2024.101051
- **Key Contributions:** Developed and implemented machine learning algorithms for a clinical decision support system, achieving an F1 score of 0.84. Optimized feature selection to identify the top 3 features, and evaluated model reliability through accuracy and correlation analysis.