Thanya Mysore Santhosh

857-339-8544 | msthanya8@gmail.com | LinkedIn | GitHub | Portfolio

Education

Northeastern University, *Khoury College of Computer Sciences* — Boston, MA
Master of Science in Data Science — *May 2026*

Vidyavardhaka College of Engineering — India
Bachelor of Engineering in Computer Science — *Aug. 2023*

Experience

Research Assistant, *Clinical Dialogue-to-Note Generation* — Sep. 2025 – Present
Northeastern University — *Boston, MA*

- Apply NLP and LLM-based techniques to convert doctor–patient dialogues into SOAP-format clinical notes, improving structure, factual accuracy, and section relevance
- Evaluate LLM-generated clinical notes for hallucinations, omissions, and logical consistency, supporting reliability analysis in healthcare NLP research

Head Teaching Assistant, *Intermediate Python for Data Science* — May 2025 – Present
Northeastern University — *Boston, MA*

- Managed and mentored 15 TAs supporting 260+ students across Python OOP, debugging, and ML coursework, ensuring smooth delivery of labs, grading, and office hours
- Optimized grading and student support workflows, reducing query backlog by 30% and improving ML project scores by 15% through targeted guidance

Data Science Intern — Jul. 2023 – Feb. 2024
66 Degrees — *India*

- Analyzed customer and marketing datasets using Python, SQL, BigQuery, and Snowflake on GCP, reducing manual analysis time by 25%
- Contributed to customer segmentation and churn prediction models using BigQuery ML and Python, improving campaign targeting effectiveness by 20%
- Built and maintained Looker dashboards on BigQuery to track churn rate, segment distribution, and campaign performance for weekly stakeholder reviews
- Collaborated with senior team members to review data workflows and model evaluation metrics, reducing recurring reporting effort by 10 hours/week

Projects

Breast Cancer Risk Modelling and Clinical Decision Support — Jan. 2026 – Present

- Develop ML models to predict benign vs. malignant tumors with calibrated probabilities for clinical decision support
- Build risk stratification pipelines to categorize patients into low, medium, and high risk groups
- Evaluate model performance and interpretability to ensure reliability in healthcare prediction tasks

Automated Due Diligence & Market Intelligence Agent — Sep. 2025 – Dec. 2025

- Built a multi-agent RAG-based GenAI system to retrieve, analyze, and synthesize evidence-backed company insights from SEC filings, news, and Wikipedia using LangGraph
- Developed a retrieval pipeline with Airflow, Docker, Qdrant, and PostgreSQL to chunk, validate, and rank large-scale documents
- Implemented CI/CD using GitHub Actions to deploy and monitor FastAPI services on Kubernetes and Streamlit on Cloud Run

Technical Skills

Languages: Python, R, C++, Java, JavaScript, SQL, NoSQL, Bash, HTML/CSS
ML & Data: PyTorch, TensorFlow, Keras, scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, Streamlit, Looker
Tools: Docker, Kubernetes, Apache Airflow, GitHub Actions, Snowflake, BigQuery, FastAPI, LangGraph