THANYA MYSORE SANTHOSH

857 339 8544 | mysoresanthosh.th@northeastern.edu| LinkedIn | Github

EDUCATION

Northeastern University, *Khoury College of Computer Sciences*   May 2026
Master of Science in Data Science
Courses: *Data Management and Processing, Algorithms, NLP, ML, Data Mining, MLOps*

Vidyavardhaka College of Engineering   Aug 2023
Bachelor of Engineering in Computer Science
Courses: *AI &ML, Mathematics, Data Structures & Algorithms, Probability & Statistics, Python Programming, SQL*

TECHNICAL SKILLS

Languages: Python, R, C++, Java, Javascript, SQL, NoSQL, JSON, HTML/CSS, Bash/Shell scripting
Data Science/Machine Learning: PyTorch, Tensorflow, Keras, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, Streamlit
Frameworks/Tools: GitHub Actions, Jenkins, Docker, Kubernetes, Snowflake, Google BigQuery, Apache Airflow, FastAPI, LangGraph
Technologies: ML, Deep Learning, Natural Language Processing, LLMs, Data Analysis & Visualization, Vector DBs (Qdrant)

WORK EXPERIENCE

RESEARCH ASSISTANT, *LLM-Based Depression Cause Analysis from Twitter Data*   Sept 2025 - Present
- Created a **Twitter API**–based pipeline to collect and preprocess **15,000+ mental-health–related tweets** for depression analysis
- Designed an **NLP** classification pipeline using **RoBERTa** to detect depression-related tweets and **LLaMA 3.3 70B** for **multi-label cause classification** across social, economic, and psychological factors
- Applied **BERTopic** to extract latent themes beyond predefined categories and analyze patterns in community-level mental-health discourse

HEAD TEACHING ASSISTANT, *Intermediate Python Programming for Data Science*   May 2025 – Present
- Managed and mentored **15 TAs**, coordinating grading, labs, and office hours for **260+** students on **Python OOP, debugging,** and **machine learning** concepts
- Monitored grading of assignments, quizzes, and online discussions on **classification, clustering, regression,** and **algorithmic problem solving,** cutting query backlog by **30%**
- Guided students and TAs in applying ML workflows including **data preprocessing, model training,** and **evaluation,** improving average student project scores by **15%**

DATA SCIENCE INTERN, *66 degrees*   Jul 2023 - Feb 2024
- Built scalable data pipelines using **Google BigQuery**, **Matillion**, and **Snowflake** on **GCP**, automating **data ingestion** from **10+** sources and reducing **data processing** time by **30%**
- Designed customer segmentation and churn modeling solutions with **Python**, **SQL**, and **BigQuery ML**, incorporating market data to improve campaign targeting accuracy by **20%**
- Deployed ML models and orchestrated **ETL workflows** with **Apache Airflow**, **Jenkins**, and **Docker**, streamlining cloud migration and saving **15** hours/week

PROJECTS

AUTOMATED DUE DILIGENCE & MARKET INTELLIGENCE AGENT
- Built a **multi-agent RAG GenAI system** to retrieve, analyze, and synthesize evidence-based company insights from **SEC filings**, **news**, and **Wikipedia** using **LangGraph**
- Developed a retrieval pipeline on **GCP** that chunks, validates, and ranks data using **Airflow**, **Docker**, **Qdrant**, and **PostgreSQL**
- Implemented **CI/CD** via **GitHub Actions** to deploy, monitor, and rollback a production system with **FastAPI** on **Kubernetes** and **Streamlit** on **Cloud Run**

LLM-ASSISTED NAMED ENTITY RECOGNITION FOR JOB DESCRIPTION INTELLIGENCE
- Developed a **BiLSTM-CRF** neural networks model with **GloVe** embeddings to extract 6 key entities from **10K+** LinkedIn job descriptions, achieving **92%** validation accuracy
- Automated annotation using **Mistral API** with **prompt engineering**, reducing labeling time by **70%** using **Generative AI** , and deployed a **Streamlit** app for real-time entity prediction

CYBERSECURITY THREAT DETECTION IN TELESURGERY SYSTEMS
- Constructed **ML** pipelines for binary threat detection, multi-class classification, and anomaly prediction, achieving over **90%** performance using **ensemble models**
- Applied **EDA**, **feature engineering**, and **SHAP**-based interpretability, achieving a **15%** reduction in false positives for threat monitoring

CERTIFICATIONS

- **GeeksforGeeks**: Complete ML & Data Science Program