NEHA JOSHI

Texas, USA | 979-344-9699 | nehayj100@tamu.edu | Portfolio | GitHub | LinkedIn | YouTube

SUMMARY

Passionate Data Scientist with 3 year's experience, currently specializing in ML at Texas A&M University. Adept in Machine Learning modeling and deployment, Large Language models and Computer Vision. Striving to contribute efficient and sustainable solutions in the industry that transform human life. Love social work and have been freelancing for an NGO since the last 2 years.

EDUCATION

Texas A&M University, College Station, Texas, USA — Aug 2023 - May 2025
Master of Science in Computer Engineering (Data Science and ML) — **CGPA: 3.9 / 4.0**
***Coursework**: Machine Learning, Programming Large Language Models, AI, Deep Learning, Software Engineering*

Visvesvaraya National Institute of Technology, Nagpur, India — July 2016 - June 2020
Bachelor of Technology in Electronics and Communication Engineering — **CGPA: 9.18 /10**
Department Rank: 2; Received Convocation Gold Medal and Academic Excellence Prize

SKILLS

Programming Languages : Python, R, SQL, C/ C++ **Databases:** MySQL, MongoDB **Cloud Platforms :** AWS
Machine Learning/AI : Deep Learning, Large Language Models, RAG, Transformers, Computer Vision
Libraries : PyTorch , Hugging Face, PySpark, Pandas, Scikit Learn, MLflow, TensorFlow, Langchain
Miscellaneous : Git, CI/CD/CT, Docker, AWS Sagemaker, GraphRAGs, PowerBI, Rest API, Streamlit

WORK EXPERIENCE

***ST*MI Lab, Texas A&M University** — *March 2024 – Aug 2024*
Machine Learning Intern | *PyTorch, Python, Computer Vision, Transformers*

- Predicted glucose spikes in diabetes patients using multi-modal data that included meal images, fitbit readings and labtest results.
- Achieved 85 % accuracy in the designed ensemble-based meal calorie prediction model pretrained on Recipe1M (13 M+ images) and fine-tuned on collected chipotle meal data for 30 subjects.
- Improved this accuracy by 10% through the designed process that used images with depth maps and pixel level annotations.
- Got 4% focal loss improvement using self-supervised model architectures (Meta's DINO and SimCLR) coded from scratch.

AspectRatio — *July 2020 - July 2023*
Senior Data Scientist | *Python, Skicit-learn, Machine Learning, LLMs, AWS, SQL, ARIMA, HWES, fbProphet, Leadership*

- Achieved ~5x reduction in analysis duration by deployment of an end-to-end ML pipeline consisting of feature extraction, pre-processing, model building, and recommendation generation for optimal product sales timing.
- Identified 1M+ new customers by conducting in-depth feature attribution analyses to understand pivotal characteristics that influence purchase of a product.
- Alleviated 60% product sales backlog stemming from the COVID-19 pandemic utilizing 4+ Time Series Forecasting models.
- Presented 20+ technical data stories to cross functional teams and directors, solving their critical business problems.

MAJOR PROJECTS

MindMend – Chatbot for a Mental Health NGO *[GitHub]* — *Aug 2024 - Nov 2024*

- Engineered a local, secure AI-powered chatbot to be used by 0.1 M + users leveraging Retrieval Augmented Generation (RAG) to provide custom mental health support in absence of a therapist.
- Performed RAG, GraphRag, LightRAG using diverse proprietary data sources like 50+ blogs, 25 Podcast, 50+ event videos.
- Achieved local token generation rate of 12 tps and evaluation rating of 3.9/5 for 101 mental health questions utilizing Qwen2, Llama 3.2 3B running locally on Macbook Air M2.
- Implemented SOS calling feature that runs in 0.34secs of user showing self-harm tendencies. Testing with 25 users currently.

Vedanta – My LLM based Personal Assistant *[GitHub]* — *Jun 2024 - Jul 2024*

- Built a LLM based local personal assistant which can schedule meetings, handle conflicts, send emails, search+summarize the web and ask me questions where the decision is not certain.
- Tuned and compared Llama 3.1 8B, Llama 3.2 3B and Mistral 7B for best performance.

Multi-Modal Deep Learning for Image Classification *[GitHub]* — *Mar 2024 – May 2024*

- Deployed a 99.8% accurate multi-modal CNN (for image) and Recurrent Neural Network (for audio) based model to production for real time digits classification.
- Achieved 1% and 9% improvement in accuracy respectively compared to individual models.

PhD Annual Review System for TAMU CSCE Department *[GitHub]* — *Aug 2023 - Dec 2023*

- Engineered a Ruby on Rails review system for 80+ professors to track and rate the progress of PhD candidates in TAMU CSE.
- Implemented features like creating student profiles, due date alerts, award declarations, flagging students with unfulfilled requirements etc. Successfully deployed the application on Heroku and is being used since January 2024.