NEHA JOSHI

Texas, USA | 979-344-9699 | nehayj100@tamu.edu | [Portfolio](#) | [GitHub](#) | [LinkedIn](#) | [YouTube](#)

SUMMARY

Passionate Data/ML Scientist with 3 years' experience, currently specializing in Data Science at Texas A&M University. Adept in Machine Learning modeling and deployment, ETL pipelines, Data Analytics, LLMs and Computer Vision. Striving to contribute efficient and sustainable solutions in the industry that transform life of billions of people around the globe.

EDUCATION

Texas A&M University, College Station, Texas | *MS in Data Science*   *Aug 2023 - May 2025*
Coursework: Machine Learning, Programming LLMs, AI, Deep Learning, Software Engineering.   **GPA: 3.9 / 4.0**

Visvesvaraya National Institute of Technology, India | *BTech in Electronics Engineering*   *July 2016 - June 2020*
Achievements: Academic Excellence Prize, Convocation Gold Medal   **CGPA: 9.18 /10**

SKILLS

Programming Languages : Python, R, SQL, C/ C++   **Databases:** MySQL, MongoDB
Libraries : PyTorch , Hugging Face, PySpark, Pandas, Scikit Learn, TensorFlow, Langchain
MLOps : GitHub Actions, Docker, AWS EC2, ECR, S3, CI/CD, DVC, ML Flow, Dagshub, Airflow, Prometheus, Grafana
Miscellaneous : Linux, Git, ETL Pipelines, RAGs, REST APIs, Agentic AI, NLP, Transformers, Distributed Training

WORK EXPERIENCE

STMI Lab, Texas A&M University   *Mar 2024 - Aug 2024*
Machine Learning Intern | *PyTorch, Python, Computer Vision, Transformers*

- Collaborated with engineers, researchers and doctors to predict glucose spikes in diabetes patients using multi-modal data including meal images, fit-bit readings and lab-test results.
- Attained 85 % accuracy in designed pipeline of ensemble-based meal calorie prediction model pre-trained on Recipe1M (13 M+ images) and fine-tuned on collected chipotle meal data for 30 subjects.
- Improved accuracy by 5% leveraging self-supervised model architectures (Meta's DINO and SimCLR) coded from scratch.

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

- Accomplished 32 hours 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 stocking.
- Initiated and led the identification of 1M+ new customers by conducting in-depth feature attribution analyses to understand pivotal characteristics influencing purchase of a product.
- Alleviated 60% product sales backlog stemming from the COVID-19 pandemic utilizing 4+ Time Series Forecasting models.
- Built MERN based interview assistant application suggesting custom questions to more than 50 interviewers in the firm.
- Effectively handled 10+ stakeholders presenting numerous data stories to help them make intelligent business decisions.

MAJOR PROJECTS

MindMend – Chatbot for a Mental Health NGO *[GitHub]*   *Aug 2024 - Nov 2024*
- Engineered a local, secure AI-powered chatbot to be operated 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 leveraging diverse proprietary data sources: 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 employing Qwen2, Llama 3.2 3B running locally on Macbook Air M2.
- Implemented SOS calling feature running in 0.34secs of user showing self-harm tendencies. Testing with 25 users currently.

Multi-Modal Deep Learning for Image + Audio 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. Employed Docker, AWS ECR, EC2 and GitHub actions to build CI/CD pipelines.
- Attained 1% and 9% improvement in accuracy respectively compared to individual models.

PhD Annual Review System for TAMU CSCE Department *[GitHub]*   *Aug 2023 - Dec 2023*
- Deployed a Ruby on Rails review system on Heroku for 80+ professors to track the progress of PhD candidates in TAMU CSE.
- Implemented features including creating student profiles, due date alerts, award declarations, flagging students etc.

Human Activity Classification Using Designed Wrist Band | VNIT, Nagpur   *Jul 2019 - Jun 2021*
- Classified diverse human activities utilizing a specialized state of the art Wrist Band. Create an end-to-end system working on real time data to perform classification.
- Employed a range of algorithms like Trees, XGBoost finalizing Dynamic Time Warping. Classification accuracy: 98.5%.