ASHWINI MURALIDHARAN

https://ashwiinii.github.io/ | linkedin.com/in/ashwini-muralidharan/ | amurali5@ncsu.edu | +1(669) 278-8612

SUMMARY

A graduate of **North Carolina State University** actively pursuing full-time opportunities as a **Software Engineer**. Proficient in **Python, R, SQL**, and skilled in frameworks like **Scikit-learn, TensorFlow, PyTorch, PySpark, SQLite and LangChain**, among others. Specialized in **Computational Intelligence** and adept at applying sophisticated data analytics, machine learning algorithms and software practices across multiple domains.

EDUCATION

North Carolina State University *Raleigh, NC, United States*
Masters in Electrical Engineering: GPA: 3.83/4.0 *Aug 2022 - May 2024*

Academic Achievements: Recipient of the **Graduate Student Support Plan (GSSP)**, a highly competitive support package providing standard tuition coverage, in recognition of academic excellence.

Relevant Coursework: *Neural Networks, Topics in Data Science, Automated Learning and Data Analysis, Digital Imaging Systems, Internet Protocols, Cloud Computing, Computer Vision*

SKILLS

- **Languages:** Python, SQL, Matlab, R
- **Frameworks:** PyTorch, TensorFlow, Scikit-learn, LangChain, Transformers, FastAPI, Flask, PySpark
- **Tools:** Tableau, Numpy, Pandas, OpenCV, AWS, Docker, Kubernetes, DBeaver, Streamlit, SQLite, Git

WORK EXPERIENCE

Union Bank of Switzerland (UBS) *New York, USA*
AI Engineer *Sep 2024 – Present*

- **Skills: Large Language Models (LLM) | SQL | LangChain | LLaMA | NLP | Fine-tuning | Python3 | Text-to-SQL | LoRA | PEFT**
- Developing an **SQL co-pilot chatbot** by curating a custom SQL dataset, automating manual SQL tasks in Python, and vectorizing the dataset, leading to a ***58%* increase in throughput** and ***45%* cost savings**.
- Contributed to the development of a **LLaMA**-powered custom **SQL co-pilot** tailored to financial professionals, achieving a ***76%*** improvement in response relevance and reduced latency over the base LLaMA model through fine-tuning techniques, including **LoRA, QLoRA, and PEFT**.

Department of Electrical and Computer Engineering *North Carolina State University, Raleigh, USA*
Natural Language Processing Engineer *Jun 2024 - Sep 2024*

- **Skills: Natural Language Processing | Retrieval-Augmented Generation (RAG) | Python3 | PyTorch | BERT | LangChain | Llama**
- Engineered a **Retrieval-Augmented Generation (RAG)** system using LLMs to perform semantic analysis of application resumes, extracting and **summarizing key achievements of applicants** with the ECE department at NCSU, leading to nearly **67% cost savings** across the department.
- Executed extensive **data preprocessing** and **annotation pipelines**, leveraging natural language processing techniques to prepare training datasets.

The Vazquez Research Group *North Carolina State University, Raleigh, USA*
Biomedical Deep Learning Engineer *Jun 2023 - Dec 2023*

- **Skills: Biomedical Signal Processing | Deep Learning | Time-series data | PyTorch | Transformers | LSTM | SciPy**
- Developed and integrated **biomedical signal processing pipeline** for cuff-less blood pressure estimation using **ECG signals**. Implemented **filtering, segmentation, hand-crafted feature extraction, data augmentation.** Developed Deep Learning algorithms using LSTMs and Transformer technologies to automate blood pressure estimation for deployment on mobile edge-devices to facilitate real-time prediction, with a ***93%* accuracy.**

Native Nibbles *Bengaluru, India*
Data Science Intern - Predictive Analytics *May 2021 - Jul 2022*

- **Skills: Data Analytics | Customer Analytics | Python Developer | SQL | Database Management | pandas | scikit-learn**
- Extracted and cleaned customer and sales data for savories and snacks, applying **DBSCAN** for clustering to enhance customer behavior predictions. Developed tailored models for each cluster, improving accuracy and processing speed for large datasets.
- Developed and optimized a **COWRF (COA-optimized Weighted Random Forest)** model, achieving a ***39.17%* increase in processing speed** and a ***97.2%* accuracy rate**, marking a ***4.7%* improvement over previous models** to evaluate the impact of promotional activities for snacks and savories, enhancing marketing strategy effectiveness for the products.

SELECTED PROJECTS

Real-time Stress Classification using Deep Learning [code-link]
Biomedical Signal Processing | Deep Learning | Python3 | Healthcare | PyTorch | SciPy

- Developed a **real-time stress monitoring system** for drivers using ***VGGNet architecture***, that classifies ECG signals collected real-time from a sensor. The **signal processing, feature extraction, training and testing of the *VGGNet architecture*** were personally carried out by me, resulting in an ***82.45%* training accuracy**.
- Implemented a system that integrates a real-time ECG sensor with a **Jetson Nano** for inference using the trained *VGGNet architecture* with the results being conveyed through a user-friendly Flask application in real-time.

Self-supervised Image Classification [code-link]
Image Processing | Self-supervised Learning | Python3 | PyTorch

- Developed a **self-supervised *SimCLR model*** using PyTorch Lightning, employing a ***ResNet-18 encoder*** and a two-layer ***MLP projection head***. Implemented advanced data augmentation and adjusted color jitter parameters to enhance training stability and speed on the STL10 dataset.
- Achieved a ***92.06%* test accuracy** by optimizing with **cosine annealing, SGD, and InfoNCE loss**. Implemented Logistic Regression on feature representations learned from SimCLR on the CIFAR10 dataset to demonstrate robust transferability of learned representations, with a 81% accuracy.

AWS Chatbot [code-link]
AWS Chatbot | Cloud Computing | Deep Learning | Flan-UL2 | Kubernetes

- Architected a cloud-based chatbot using ***Flan-UL2*** model for real-time, automated customer support, ensuring 24/7 availability, high scalability, and security, and reducing operational costs.
- Implemented and managed an **AWS Elastic Kubernetes Service (EKS)** cluster to support a scalable, high-performance chatbot infrastructure, handling elastic scaling from 4 to 20 pods to maintain seamless user interactions and optimal resource utilization.