**SHRUTI BALAJI**

<https://github.com/ShrutiBalaji> || <https://www.linkedin.com/in/shrutibalaji17/> **||** [shrutibalaji17@gmail.com](mailto:shrutibalaji17@gmail.com)

(508) 834-2626 || West Point, PA

**EDUCATION**

**Master of Science, Data Science**

University of Massachusetts, Dartmouth, USA  **(2023 - 2025)**

**Bachelor of Technology, Bioinformatics**

SASTRA Deemed University, Thanjavur, India **(2019 - 2023)**

**PROFESSIONAL EXPERIENCE**

**AI ENGINEER Intern (July 2025 – September 2025)**

***ARMA AI Labs INC, Remote***

* Working on audio model pipeline using a LLM trained on billions of parameters. Finetuning using **LORA** and **PEFT** to train a model on TTS pipeline with an emotional layer that needs to be evaluated on all types of accents.
* Leveraging CUDA GPU acceleration with PyTorch to efficiently run and test large-scale models comprising billions of parameters.

**DATA SCIENCE LAB RESEARCHER (October 2023 – May 2025)**

***Multi-scale Medical Robotics Lab at UMassD***

* Designed a computer vision model leveraging the **Faster R-CNN** architecture to detect tumors and perform breast cancer segmentation on DICOM images of the MG modality. Extracted regions of interest from each image using bounding boxes based on annotation files. Extended the project to incorporate training a multimodal fusion algorithm utilizing multiple image modalities.

**DATA ANALYTICS INTERN (June 2024 – August 2024)**

***Oceantic Network, Boston, MA***

* Performed data cleaning, manipulation and analysis, data visualization on past trends to give business insights for key performance indicators such as percentage increase in job numbers over the past few years, percentage increase in job numbers, trends in the market, percentage increase in skilled and unskilled laborers, etc.
* Conducted time series forecasting and predictive analysison employment trends in the offshore wind industry using **ARIMA models,** predicting a 7% increase in job opportunities by 2025 based on seasonal patterns and government policy influence and provided a detailed stakeholder documentation report and business intelligence dashboards using Tableau.

**SOFTWARE DEVELOPER TRAINEE/INTERN (February 2023 – June 2023)**

***Psiog Digital, Chennai, India***

* sq

**DATA ANALYTICS INTERN (February 2022 – March 2022)**

***Rela Multi-specialty Hospital, Chennai, India***

* Designed and developed an **interactive dashboard** leveraging call center data of the hospital to provide actionable business insights, including key performance indicators (KPIs) such as call volume trends, average handling time, most in-demand department, availability of doctors in each, average time taken to find appointment slots, etc. The dashboard enabled real-time monitoring, data-driven decision-making, and operational efficiency improvements across customer support teams.
* Integrated the tabular data format into **Tableau**, created calculated fields, chose the visualization type (line chart or bar chart, heat map, etc.), organized metrics to visualize to provide data insights. Also made use of **R programming** language for simpler statistical modeling and business requirements.

**MACHINE LEARNING INTERN (February 2021 – February 2023)**

***Computational Lab at SASTRA Deemed University, Remote***

1. **Breast Cancer Stage Classification and Segmentation**

* Developed and trained deep learning models including a 10-layer **CNN, ResNet50, and Faster R-CNN** for multi-class cancer stage classification and tumor detection on DICOM mammograms from TCGA online database.
* Achieved 94.6% accuracy with CNN and optimized model performance via parallelized training on 64-core Carnie (UMassD supercomputer), reducing training time by 15%. Engineered region-based segmentation using bounding boxes derived from annotation files for precise tumor localization. Extended the pipeline by integrating multimodal fusion techniques to enhance diagnostic accuracy across multiple imaging modalities.

1. **Screening of compounds for inhibitors**

* Developed a **Random Forest** binary classification model to screen for inhibitors and non-inhibitors.
* Performed data balancing with **SMOTE**, feature selection, and comparison of dimensionality reduction using **PCA and LDA,** achieving 87% accuracy with PCA. The model efficiently predicted the inhibitors for in-silico analysis from a database containing lakhs of compounds, saving immense time and screening the compounds 90x times increase in screening efficiency.

**PERSONAL PROJECTS ON GEN AI AND LLMs**

**Conversational Chatbot for Course Recommendations (March 2025 – Present)**

* Developing a real-time conversational chatbot for the UMass Dartmouth website that accepts student resumes, analyzes their skills and strengths, identifies skill gaps, and recommends relevant university courses to address those gaps.
* The chatbot leverages the **LangChain** frameworkforintegration, **OpenAI embedding, FAISS vector database, GPT-4 for LLM as Retrieval-Augmented Generation (RAG) stack** to build an intelligent, personalized course recommendation system.

**Financial data streaming and processing pipeline (Sept 2024 – Dec 2024)**

* Designed a **data ingestion and processing pipeline** using historical stock data and simulated a real-time streaming environment using **Kafka**, mimicking live data ingestion and processing pipelines for financial market data. Created **Python scripts** to establish a producer-consumer architecture for data ingestion and processing through Kafka brokers.
* Integrated **IAM roles** for security management and implemented data storage using S3 buckets for downstream analytics, integrated **ETL workflow** by extracting JSON data format in **Amazon Glue**, transformed into structured format and loaded back into S3. Used **Athena** to make necessary data insights and analyze the data using **SQL** queries.

**Building a Nano LLM for text generation – GenAI and NLP (January 2024 - May 2024)**

* Engineered an **n-gram Nano Language Model** using **transformers** with character tokenization, which learns from the user-provided text and generates subsequent words similar to the provided data. The retained model memory is pickled and transferred to be used later for text generation based on the trained data.

**CERTIFICATION**

Data Science course offered by IIT Madras, Technical Support Fundamentals offered by Google, Bit and Bytes of Computer Networking offered by Google, Operating System offered by Google, Web Programming by EdX, Python Boot Camp offered by Udemy.

**SKILLS**

**Programming Languages:** Python (NumPy, Pandas, Streamlit, Django), Assesmbly, R (Shiny), Rust (beginner), Perl, C, C++, C#

**AI / ML Concepts:**

* Machine Learning: Scikit-learn, Predictive Modeling (Time Series Forecasting, Regressions)
* Deep Learning: TensorFlow, PyTorch, TorchVision, Hugging Face Transformers
* Advanced AI: Fine-tuning LLMs, Retrieval Augmented Generation (RAG), Transformers, Topic Modeling (LDA, BERT), NLP, Computer Vision (YOLO, CNN, R-CNN, Faster R-CNN), Audio Processing (TTS & voice cloning with LLaSA 3B)

**Agentic AI:** Multi-agent systems and orchestration using LangChain, AutoGen, CrewAI, n8n, Cursor; workflow automation; vector databases for RAG (FAISS, Pinecone, Weaviate, Chroma); tool integration and agent coordination

**Cloud & Deployment:** AWS (Kafka, Glue, Athena, EMR Serverless, QuickSight, SageMaker), Docker

**Databases / DBMS:** MySQL, MSSQL, PostgreSQL, MongoDB, SQL (advanced querying, schema design, ETL workflows)

**Data Visualization:** Tableau, Plotly, Matplotlib, Seaborn, ggplot2, Power BI, Power Query

**Core CS & Development:** CUDA, High-Performance Computing, Linear Algebra, Data Structures & Algorithms, Statistics

**Web Development:** HTML, CSS, JavaScript, ReactJS, PHP

**Version Control:** Git/GitHub for collaborative development