Sriphani Vardhan Bellamkonda

bellbytes.dev 3477448482 sriphani345v@gmail.com 1143, 67th Brooklyn, New York, 11219

Education

New York University **Sept. 2023 - Present**
Master of Science in Computer Science; GPA: 4.0/4.0 *Manhattan, NY*

National Institute of Technology Warangal (NITW) **Aug. 2018 – May 2022**
Bachelor of Technology in Computer Science and Engineering; GPA: 7.2/10.0 *Telangana, India*

Technical Skills & Relevant Courses

Languages: C, C++ , Python , Java ,Javascript, Typescript
AI ML Frameworks: numpy, pandas, scikit-learn, PyTorch, Tensorflow, AWS Sagemaker
Databases: MySQL, SQL, AWS DynamoDB, SQLAlchemy, MongoDB, NoSQL
Big Data & MLOps: Hadoop, Spark, Kubernetes, Docker, CI/CD, HDFS, MapReduce, AWS
Web Technologies: HTML, CSS, Django, Flask, Spring Boot MVC (Java), FastAPI, Node, React, Angular, API, Unix
Certifications: AWS Solutions Architect SAA-C03, Azure Fundamentals AZ900
Coursework: Data Science, Linear Algebra, Machine Learning, Statistics, Software Engineering, Databases, Big Data

Work Experience

Machine Learning Engineer — New York University — Manhattan, New York **March 2024 - Present**

- Contributed to the DARPA funded CCU project to identify cross-cultural misunderstandings in dialogue.
- Engineered Multilingual Speech Recognition and Diarization using Whisper, implemented with PyTorch and Hugging Face Transformers, processing over 9 terabytes of multilingual audio/video data on Linux High-Performance Computing Clusters.
- Designed and optimized end-to-end ML pipelines for data preprocessing, model inference, and evaluation, utilizing Python, Numpy, and Pandas, orchestrated via SLURM, resulting in a competitive Word Error Rate (WER) of approximately 10% in English.
- Programmed shell scripts to crawl and backup 10 TB and files and directories with deduplication strategy.

Software Engineer — Wells Fargo — Hyderabad, India **July 2022 - July 2023**

- Designed scalable architecture for an enterprise system with 10+ modules, employing Java Spring Boot and Angular Typescript.
- Attained a code coverage of over 95% through unit-testing with JUnit and Mockito, ensuring software robustness and reliability
- Migrated a legacy database from MS Access to SQL Server and developed an automated query translator, reducing manual adjustments by 98%.
- Developed and optimized 50+ backend APIs for server-side pagination, filtering, search, and CRUD operations, and implemented ETL processes for Oracle databases, leading to an 80% improvement in query performance.

Software Engineer Intern — Wells Fargo — Hyderabad, India **May 2021 - July 2021**

- Architected a scalable full-stack Deep Learning solution for credit risk modelling with 5 classification models and performed Exploratory Data Analysis and Preprocessing on the Dataset along with model training on GCP.
- Developed dashboard monitoring with Flask (Python) & Firebase(NoSQL) optimizing and retrieval speed less than 500ms
- Implemented user authentication using OAuth 2.0, token-based authorization mechanisms with JWT, and caching which boosted performance by 70%.

Undergraduate Researcher — NIT, Warangal — Warangal, India **Sept. 2021 - May 2022**

- Designed a novel mini-batch sampling method that leveraged feedback from previous batches, enhancing model and performance.
- Developed custom machine learning training scripts using TensorFlow and scikit-learn, improving model accuracy by 1-4%.
- Experimented with Logistic Regression, Neural Networks, and CNNs on 4 datasets for classification tasks, refining models for improved performance.

Projects & Publications

(ACL) NLP Transformer models on Low Resource Language | *PyTorch, Web Scraping*

- Curated a Telugu Humor Dataset and fine-tuned Multilingual BERT models for sentiment analysis on an A100 GPU, achieving an F1 score of 0.82, and published in the ACL 2022 Workshop using XLM-RoBERTa.

Machine Learning Pipeline for Text Summarization with Sagemaker (MLOPS) | *Python, Transformers, HuggingFace, AWS*

- Orchestrated the data ingestion, validation, transformation, model training, and evaluation by 2 methods.
- Method 1: Containerized the app with Docker and AWS EBS, served the inference and training endpoints on AWS EC2 with the backend by FastAPI and NGNIX, performed Automated deployments and CI/CD with Github Actions and version control.
- Method 2: Preprocessed raw data and stored in AWS S3 bucket, applied Lambda to trigger a training job using Hugging Face Deep Learning Container and EC2, then served via AWS Sagemaker endpoint while managing model versions with model registry.

MongoDB Flask app with Kubernetes | *AWS EKS, Kubernetes, Flask, MongoDB*

- Built a Flask (Python) app with MongoDB, deployed in a Linux environment using Minikube and AWS EKS, and configured replication controllers with 3 replicas and a rolling update strategy for high availability.

Lightweight ResNet Architecture | *HPC, SLURM, PyTorch*

- Optimized ResNet architecture by modifying block sizes, kernels, and skip connections, achieving 96.2% accuracy on the CIFAR-10 dataset with fewer than 5 million parameters, outperforming ResNet-18's 86% accuracy.

Big Data Analytics on NYC Taxi Data | *PySpark, Hadoop Map Reduce*

- Processed large-scale NYC Taxi data on HDFS using Hadoop MapReduce and PySpark, performing analytics with custom parallel queries, boosting processing efficiency by over 85%.