Jinam Shah

Raleigh, NC | (919)922-8481| jinams47@gmail.com | github.com/jinamshah | linkedin.com/in/jinamshah

Experienced Machine Learning Engineer with three years of expertise in developing ML, NLP, and Big Data applications. Proficient in AWS and GCP. Skilled in designing scalable systems and leading innovative projects. Developed an NLP product that classifies documents into 1500 categories and a big data platform managing 1.5 TB of data daily, accumulating over 900 TB in a data lake.

WORK EXPERIENCE

Software Engineer, North Carolina State University | June 2023 – Present

- Leading projects funded by the US Department of Agriculture.
- Created a **data pipeline** for processing drone images, handling over 100GB of data per flight.
- Utilized photogrammetry and **ML** to extract insights from drone images for **resilient agriculture practices**.

Machine Learning Engineer, Cactus Communications Inc | May 2022 – Nov 2022

- Managed large-scale **pattern recognition** on **250 TB** of raw data.
- Successfully designed and implemented a scalable **ML product** in under a week with **zero downtime**.
- Developed a **big data platform** that ingests over 1.5TB daily and generates approximately 8TB weekly, managing **900TB in the data lake**.

Senior Software Engineer, Cactus Communications Inc | June 2020 – July 2021

- Designed and implemented a highly efficient **data processing pipeline** for a machine learning product, utilizing **24K CPU cores and 48Tb RAM** to generate over 4.5Tb of data in under 2.5 hours while achieving **1/5th the cost proposed by the AWS Big Data Team**.
- Designed and implemented a serverless, highly available API that efficiently handles **~100K long-running requests** with a **sub-second SLA**.

Software Engineer, Cactus Communications Inc | June 2018 – May 2020

- Designed and developed **image recognition** products, leading the charge to create a **new business vertical** for the company.
- Spearheaded the **architecture planning** and implementation of various ML/NLP and big data products, effectively **bridging the gap** between business and technical teams.

TECHNICAL SKILLS

Programming tools: Python, Pytorch, Tensorflow, Javascript, React, Pandas, Django, Flask, Spark, SQL, Git, Hadoop, Java, Airflow, Big query, scikit-learn, docker

Domain Expertise: Machine Learning, Natural Language Processing, Image Recognition, Large Language Models, Distributed Training, Deep Learning, Big Data, Transformer-based language modeling

Cloud technologies: AWS, GCP, AWS EC2, AWS S3, AWS Lambda, AWS RedShift, AWS Kinesis, AWS API Gateway, AWS CloudFormation, AWS ECS

Design Principles: Cost-effective architecture, scalable and secure architecture, serverless architecture.

EDUCATION

Master of Science in Computer Science, North Carolina State University | August 2021 - May 2023
GPA: 4.0
Thesis: Author name disambiguation in academic records

Bachelor of Technology in Computer Science, NMIMS University | August 2014 - May 2018
GPA: 3.91
Research: Image recognition and image caption generation