Jinam Shah

Raleigh, NC | (919)922-8481| jbshah@ncsu.edu | github.com/jinamshah | linkedin.com/in/jinamshah

EDUCATION

Master of Science in Computer Science, North Carolina State University *August 2021 – May 2023*

GPA: 4.0

Relevant Courses: Artificial Intelligence-1, Neural Networks, Natural Language Processing, High-performance Machine learning and Real-time AI, Automated Learning and Data Analysis, Design and Analysis of Algorithms, Software Engineering, Algorithms for Data Guided Business Intelligence.

TECHNICAL SKILLS

Programming languages and tools:	Python, Keras, Pandas, Django, Flask, Spark, SQL, Git, C, C++
Domain Expertise:	Machine Learning, Deep Learning, Natural Language Processing, Image Processing, Distributed Training, Big Data
Cloud technologies:	AWS, GCP, EC2, ECS, S3, Lambda, RedShift, Kinesis, API Gateway, CloudFormation
Design Principles:	Cost-effective, secure, and reliable architecture design, Serverless architecture, Reproducible infrastructure

WORK EXPERIENCE

Machine Learning Intern, Cactus Communications, Princeton, USA *May 2022 –Present*

- Working with the director of data engineering on various **big data** problems and **pattern recognition** problems based on the company's datasets.
- Own the **Machine learning track** for disambiguation of records in a data lake.
- Architected a truly serverless highly available API that orchestrates **~100K long running requests** with **sub-second SLA**.

Senior Software Engineer, Cactus Communications, Mumbai, India *June 2020 – July 2021*

- Led the **architecture planning** and implementation of various products in ML/NLP and BigData, **bridging the gap** between the business and tech teams.
- Reported and **worked with AWS S3 team** on fixing bug on prefix throughput.
- Designed and developed a pipeline to process data for a Machine Learning use case with **cumulative 24K vCPU, 48Tb Memory**. Generated over ~4.5Tb of data in under 2.5 hours at **1/5th the proposed cost from AWS Big Data Team**.
- Designed and implemented a **BigData platform** that ingests over 1.5Tb/week and generates over 4.5Tb/week. It manages over ~250Tb of data in the data lake.
- Setup the best practices and implementation guidelines for the team to operate on Cloud across AWS, Azure and GCP. **Templatized infrastructure.**

RESEARCH WORK AND PROFESSIONAL PROJECTS

Author Name Disambiguation (Ongoing research)

- Working on performing **bias-free** author name disambiguation for research manuscripts and improving the state-of-the-art.
- Based on the research done by **AllenAI** on the same topic.

Transformer-based Document Classification

- Ensemble of DL and ML models (based on **BERT**) performing document classification for over an unprecedented **1500 classes** deployed using **serverless** architecture.
- Saved the organization around **$1M per annum** and reduced the TAT for the service from **8 hours to under 2 minutes** (down by 99.6%).

Serverless Image Recognition

- Image recognition software for determining **ethical compliance of images** added in research papers.
- Achieved state-of-the-art performance (**99.8% accuracy**) and deployed using **serverless** architecture.

Automated Language Correction

- Worked on the software engineering and "explainable AI" portion behind an NLP software that focused on automated grammar correction.
- This product led the efforts of the business to create a new vertical for generating revenue of around $1M.

ACTIVITIES

Open source:	Contributor of AllenAI's S2AND and Specter repositories, consistent contributor through HacktoberFest.
Hackathons:	Top 25 percentile in Reply Code hackathon.
Scholarship:	AWS Machine learning engineer scholarship by Udacity.
Volunteer:	AI4Good foundation, working on solving United Nations' Sustainable Development Goals.