

Nikhil Naik

Austin, TX 78717 | nnaik2@stevens.edu | [LinkedIn](#) | [GitHub](#) | [Medium](#) | +1 (917)282-6968

SUMMARY

Machine Learning Engineer with 3+ years of experience in GenAI, ML, software development, and project management. Demonstrated expertise in developing Llama3 and Mistral LLMs, DL models, and architecting large scale software systems. AWS Certified Solutions Architect with deployment experience on Nvidia Triton and AWS Cloud

EDUCATION

Stevens Institute of Technology; Hoboken, New Jersey

May 2024

Masters in Machine Learning

CGPA: 3.96 / 4.0

Courses: Deep Learning, Big Data Tech, 3D & 2D CV, NLP, Knowledge Discovery and Data Mining

PROFESSIONAL EXPERIENCE

Stevens Institute of Technology; Hoboken, NJ, USA

Research Project – Workshop on Machine Translation

Jan 2024 – Present

- Representing Stevens at an International conference, contributing novel research in multilingual machine translation
- Developed finetuning pipeline for low resource data using Fairseq for MT-5, NLLB, Transformer model
- Innovated an iterative back-translation system generating synthetic data, effectively raising the BLEU score to 42.2

Gajshield Infotech Firewalls; Mumbai, India

Software Developer & Machine Learning Engineer

Sept 2018 – Mar 2021

- Developed in-house malicious email detecting system based on Scikit-learn's outlier detection model with 96% Recall improving quality and performance by 40% of the Firewall's email security system
- Developed a Technical Assistant Chatbot using RASA to help the operation support team efficiently resolve customer issues, resulting in a 30% reduction in the engineering team's ticket and workload.
- Led a 4-member team that developed two industry-scale Cloud-managed VPN and SD-WAN projects using open-source Linux libraries and MySQL. Instrumental in securing a major BPCL contract worth 6 million dollar
- Enhanced and modernized frontend of Gajshield Firewall's UI, utilizing Web2py, ExtJS, Pygal, and Chartjs
- Served as Gajshield's technical representative at stakeholder meetings, collaborating with Bharat Petroleum to architect Zero Touch Provisioning deployment of over 300 Firewalls and cloud-based management
- Optimized productivity by initiating Kanban and Agile methodologies into Gajshield's SDLC process

TECHNICAL PROJECTS

Advanced Healthcare LLM Fine-Tuning and Deployment - [Medium Article](#)

Present

- Executed parameter-efficient fine-tuning on Llama3 and Mistral MOE LLM using proprietary healthcare data
- Employed tools like LlamaFactory, FlashAttention, DeepSpeed, and Unsloth to execute quantization and LORA techniques, achieving 3x faster fine-tuning on cloud platforms such as Lambda Labs and Vast AI
- Deploying with TensorRT and Triton Inference Server to handle 30-40 concurrent asynchronous responses
- Developing a LangChain and LlamaIndex powered Agentic RAG system to enhance performance and reliability

Forex Rate Prediction – Deep Learning - [GitHub](#)

Mar 2023

- Architected a baseline RNN & LSTM neural network with In-Domain world Forex dataset using TF and Keras
- Hyper parameter tuned those models and used transfer learning fine-tuning on target data to achieve 0.45 R2 score

CitiBike Usage Forecast – Big Data ML - [GitHub](#)

Apr 2023

- Leveraged PySpark on Google Cloud Platform's Dataproc to create ETL pipeline on 40GB time-series dataset
- Utilized Spark's MLlib to train a distributed multivariate regression model achieving a 0.85 R2 score

Telemarketing Campaign Prediction - [GitHub](#)

Sept 2021

- Performed Exploratory Data Analysis to gain data insights, selected and tuned best Machine Learning algorithm
- Awarded excellence certificate for achieving an accuracy of 95% employing KNN Classifier with silhouette analysis

TECHNICAL SKILLS & ACCOMPLISHMENTS

Certifications:	AWS: Solutions Architect Associate - Cert	IIT Roorkee's Data Science - Cert
	Fundamentals of Data Mining - Cert	Algorithmic Toolbox, Coursera - Cert
Tech Stacks:	Python, C++, MLOps, Shell, JavaScript, MySQL, SQLite, MSSQL, JSON	
Frameworks:	TensorFlow, PyTorch, Keras, Fairseq, HuggingFace, LangChain, OpenCV, Spark, Django, Flask	
Other Libraries:	Scikit-Learn, Transformer, Amazon Web Services, SageMaker, S3, EC2, Lambda, GCP, Docker, Kafka, Linux (CentOS), Apache Web Servers, Full-Stack, SDLC, Nvidia-Triton, Kanban, Agile	
Publication:	Sign Language Interpreter Glove IEEE ICATE-2K18 Event ISBN: 978-93-5267-422-0	