Nikhil Naik

Austin, TX 78717 | nsnaik1996@gmail.com | LinkedIn | Portfolio | GitHub | Medium | +1 (917)282-6968

SUMMARY

Machine Learning Engineer with 3+ years of expertise in Generative AI, LLM fine-tuning, and deep learning. Led teams on projects securing multi-million-dollar contracts, and developed scalable, cloud-based solutions. Proven skills in building RAG systems, optimizing performance, and creating AI-driven products in healthcare and insurance. AWS Certified Solutions Architect with hands-on experience in deploying on Nvidia Triton Inference Server

EDUCATION

Stevens Institute of Technology; Hoboken, New Jersey **May 2024**
Masters of Science in Machine Learning CGPA: 3.96 / 4.0

PROFESSIONAL EXPERIENCE

Machine Learning Engineer – GenAI Intern - *Medium Article*
LifeSage Inc; Austin, USA **July 2024 - Present**

- Developed and fine-tuned Healthcare and Insurance LLM assistants using Llama3 and Gemma2, enabling accurate policy-specific queries and healthcare advice through Text-to-SQL AI agents with Vanna
- Achieved 25% ROUGE score improvement with a BERT-based data cleaning pipeline
- Attained 40+ concurrent request handling by architecting scalable inference web apps with FastAPI, TensorRT, and vLLM, deployed on Triton Inference Servers
- Built PDF parsing systems with AWS Textract, storing structured data in SQL, Pinecone, and Qdrant for RAG
- Delivered AI-driven POCs and demos to stakeholders in healthcare and insurance sectors

Machine Learning Engineer & Software Developer
Gajshield Infotech Firewalls; Mumbai, India **Sept 2018 – Mar 2021**

- Achieved 40% reduction in computational workload of the email security system by integrating an ML-based outlier detection model, accurately detecting malicious emails with 96% recall
- Developed a Technical Assistant Chatbot using RASA to help the operations team efficiently resolve customer issues, resulting in a 30% reduction in the engineering team's ticket and workload
- Architected and led a 4-member team to develop two cloud-based VPN and SD-WAN projects using Docker, Linux, Python and MySQL. Instrumental in securing a major tenure worth $6 million
- Experienced in understanding stakeholder requirements, designing feasible products, creating intuitive UIs and visualizations, and managing cloud-based deployments

TECHNICAL PROJECTS

Multilingual Translation System – Workshop on Machine Translation 2024 **July 2024**

- Representing Stevens at an international conference with novel multilingual machine translation research
- Developed an iterative back-translation finetuning pipeline for NLLB and MT-5 models, raising BLEU score to 59.8

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 time-series forecasting with 0.45 R2

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, Pydantic, Kafka, Linux, Apache Web Servers, Full-Stack, SDLC, Nvidia-Triton, Gradio, Agile	
Publication:	Sign Language Interpreter Glove | IEEE ICATE-2K18 Event | ISBN: 978-93-5267-422-0	