KARTHVIK SARVADE

+1(929) 308-6751 ◇ Atlanta, GA
s.karthvik@gmail.com ◇ linkedin.com/in/karthvik/ ◇ github.com/Karthvik

SUMMARY

Machine Learning Engineer with expertise in fraud detection systems, LLMs, and computer vision. Demonstrated success in model optimization, MLOps, and real-time data pipelines, reducing latency and improving performance across financial applications.

EXPERIENCE

Machine Learning Engineer July 2021 - Present
Captial One *New York, NY*

- Lead development of multimodal fraud detection systems combining graph neural networks **(GNNs)** and transformer architectures, reducing synthetic identity fraud by 32% across credit card portfolios.
- Designed a real-time feature store using **Apache Kafka** and **Redis**, reducing the inference latency for transaction risk models from 120 to 45 minutes.
- Optimized Capital One's **Eno virtual assistant** by fine-tuning a distilled **BERT** model for intent classification, improving accuracy by 18% while reducing inference costs by 40% using **AWS Inferentia**.
- Designed a **reinforcement learning** framework for dynamic adjustment of the credit limit, increasing customer satisfaction scores by 22% through personalized offers.
- Prototyped a computer vision system for document verification using **EfficientNet**, achieving 98% OCR accuracy on driver's license scans.
- MLOps Pipeline Automation: Developed Terraform modules for provisioning SageMaker pipelines, standardizing CI/CD workflows across 20+ ML teams, and cutting deployment time from 2 weeks to 3 days.

Research Assistant Sept 2019 - May 2021
New York University *New York, NY*

- Contributed to the **M²LInES** initiative, a Schmidt Futures-funded project using ML to reduce biases in climate models. Developed neural closure models for ocean turbulence parameterization in **PyTorch**, improving simulation accuracy by 18% compared to traditional physics-based methods.
- Collaborated with climate scientists and ML researchers to analyze 50TB of high-resolution climate simulation data (CESM, GFDL) using Dask and Xarray, identifying key patterns for mesoscale eddy parameterization.
- Optimized training workflows for transformer-based models on NYU's HPC clusters (GPUs), reducing convergence time by 35% via mixed-precision training and gradient checkpointing.

EDUCATION

Master of Science in Computer Science, New York University, **GPA: 4.0/4.0**
Relevant Coursework: Advanced Machine Learning, Deep Learning, Big Data and ML for Cybersecurity

Bachelor of Science in Computer Science, Jain University, **GPA: 3.8/4.0**
Relevant Coursework: Cloud Computing, Principles of Database Systems and Data Structures and Algorithm

SKILLS

Programming	Python, SQL, C, C++, Java
ML Frameworks	PyTorch, TensorFlow, Hugging Face
Cloud and MLOps	AWS SageMaker, Azure ML, GCP Vertex AI, CI/CD Pipelines
Data Technologies	Apache Kafka, Redis, Delta Lake, Dask, Xarray
ML Specialties	GNNs, Transformers, BERT, GPT-4, Reinforcement Learning, Computer Vision
Databases	PostgreSQL, MongoDB, MySQL, BigQuery