Aditya Rustagi

LinkedIn: aditya-rustagi54
Github: github.com/AdityaR-Bits
Email: aditya.rustagi54@gmail.com
Mobile: 8583199631
Website: adityar-bits.github.io/personal_profile

Professional Summary

Machine Learning Engineer with 3+ years of experience in designing, deploying, and optimizing scalable ML models in production. Skilled in MLOps, cloud-native ML platforms, and data pipelines. Proven track record of improving forecast accuracy, reducing compute costs, and mentoring teams.

Experience

- **CVS Health** Apr 2023 - Current
  Machine Learning Engineer
  - Pioneered the development of a fine-tuned LLM pipeline, containerized with Docker and serving via FastAPI, with workload orchestration enabling scalable processing of millions of product records per run
  - Directed the optimization of critical batch processing pipelines in Azure, resulting in a 98% reduction in compute time on more than 3 billion records weekly with $2.4 Mil annual cost savings
  - Designed and implemented robust MLOps frameworks, incorporating model drift detection and automated model retraining, comprehensive error handling, and centralized logging, reducing manual intervention by 30% and improving model lifecycle management. Contributed to organization wide internal tooling with these packages
  - Developing an MCP based multi-agent AI system for domain-specific question answering within the organization
  - Mentored junior engineers on scalable ML model deployment and CI/CD pipelines for cloud-based ML inference. Collaborated with business and product stakeholders to define ML requirements and align on KPI-driven goals
- **NVIDIA** June 2022 - Sept 2022
  Software Engineering Intern - RAPIDS Data Science
  - Developed CUDA-accelerated data science kernels within the RAPIDS ecosystem, achieving 6× performance improvements on large-scale tabular datasets. Optimized rendering backend operations by integrating HoloViews visualization, reducing latency and improving small-dataset rendering throughput by 2×
  - Contributed optimizations to open-source LUX and RAPIDS libraries, advancing GPU-native data visualization
- **Rootee Health** Apr 2020 - July 2021
  Machine Learning Engineer
  - Developed and deployed optimized TensorFlow models for real-time object tracking and segmentation on edge devices, leveraging TensorRT for a 50% inference speed enhancement, critical for medical diagnostic applications
  - Implemented anomaly detection models achieving a 92% reduction in faulty images being processed
  - Deployed the flagship eye illness diagnostic pipeline for the startup, demoed to investors leading to an acquisition
- **Central Electronics Engineering Research Institute** Aug 2019 - Dec 2019
  Machine Learning Engineer Intern
  - Designed a deep learning 3D CNN based custom architecture for Yoga Pose estimation published in a journal
  - Surpassed the state of the art with 99.39% accuracy and real-time performance at 20 fps using TensorRT

Projects

- **AI Document Assistant: LangGraph-Orchestrated RAG Pipeline**:
  - Built LLM-powered document assistant with LangGraph orchestration, implementing semantic search using ChromaDB embeddings and intelligent tool routing for question answering on any given document
  - Engineered RAG pipeline using FastAPI, HuggingFace embeddings, and OpenAI GPT-3.5-turbo with optimized text chunking and persistent vector storage for real-time document analysis
- **Tweet Phrase Extraction via Sentiments in NLP (Research at UC San Diego)**:
  - Context-Aware Sentiment Phrase Extraction with Transformer Models: Engineered a highly accurate, BERT-based Natural Language Processing model for fine-grained sentiment phrase extraction within a Question-Answering paradigm
  - Utilized advanced tokenization strategies with analysis of attention-weighted embeddings achieving a 0.85 Jaccard score

Education

- **Master of Science in Machine Learning and Data Science**
  University of California San Diego
- **Dual Degree - B.E in Electrical and Electronics & M.Sc in Chemistry**
  Birla Institute of Technology and Science, Pilani

Skills Summary

Languages: Python, SQL, R, C++, CUDA, GO, MATLAB
Frameworks & OS: PyTorch, TensorFlow, Linux, Windows, Mac OS
Libraries: Hugging Face, LangChain, Spark, Pandas, Numpy, Sklearn, TensorRT, SciPy, Dask, ChromaDB
CI/CD Tools: Vertex AI, GCP, Azure, Kubernetes, Kubeflow, Airflow, Jenkins, Docker, Github