Aditya Rustagi

LinkedIn: aditya-rustagi54
Github: github.com/AdityaR-Bits

Email: aditya.rustagi54@gmail.com
Mobile: 8583199631

EXPERIENCE

- **CVS Health** — Apr 2023 - Current
  Machine Learning Engineer
  - Architected and led the implementation of a real-time, scalable demand forecasting and price optimization machine learning platform, driving an 7% increase in sales across key product categories
  - Pioneered the development of a GenAI-powered retail basket optimization pipeline, leveraging fine-tuned LLMs and causal inference to deliver product pairing recommendations
  - Directed the re-platforming of critical batch processing pipelines to a serverless, event-driven Azure architecture, resulting in a 98% reduction in compute time for 100 Mil + records daily with $1.2 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
  - 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
  - Deep expertise with LLMs, transformers, LSTMS, tree models, boosting, deep learning, and ARIMA
  - Utilizing Azure, Kubernetes, Docker, Airflow, Kubeflow, Spark, SQL, Python, Snowflake, Jenkins, Dask etc

- **NVIDIA** — June 2022 - Sept 2022
  Software Engineering Intern - RAPIDS Data Science
  - Engineered 3-9x computational efficiency improvements for large datasets using CUDA and RAPIDS
  - Spearheaded HoloViews integration redesign, achieving 2x improvement in small dataset rendering times
  - Delivered key contributions to the official LUX data visualization library

- **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 in PyTorch, achieving a 92% reduction in faulty images being processed
  - Successfully built a complete AI pipeline for eye illness diagnostics for the startup, 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

- **Large Language Model based cost optimization**:
  - Generative AI-driven Cloud Cost Optimization Platform: Prototyping an LLM-powered decision engine to dynamically recommend optimal Snowflake warehouse configurations
  - The system analyzes query complexity, data volume, and historical cost patterns to minimize expenditure. Preparing for leadership pitch at CVS Health

- **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, C, C++, CUDA, GO, MATLAB
Frameworks & OS:	PyTorch, TensorFlow, LLMs, Keras, Linux, Windows, Mac OS
Libraries:	Spark, Pandas, Numpy, Sklearn, Xgboost, TensorRT, SciPy, OpenCV, REST
CI/CD Tools:	Kubernetes, Kubeflow, Airflow, Jenkins, Docker