Rohith Sri Sharan Jangam

LinkedIn | +91 8500976792 | https://rohithj.com | jangamrohith4@gmail.com | Github

Career Summary

Results-driven MLOps Engineer **with** 2+ years of combined professional and academic experience **designing and managing end-to-end machine learning pipelines. Skilled in model training, deployment, and monitoring on** GCP (Vertex AI) **and** AWS**, with strong expertise in** Python, TensorFlow, and orchestration frameworks (TFX, Kubeflow, Airflow)**. Proven track record of implementing** scalable, automated workflows **that ensure high availability, reliability, and performance of ML systems. Adept at collaborating in** Agile teams **to deliver impactful, data-driven solutions. Passionate about advancing** MLOps practices**, optimizing pipelines with** GPU acceleration**, and driving innovation in production ML environments.**

Skills

Programming & Libraries: **Python (Pandas, NumPy, TensorFlow, Scikit-learn), SQL, Bash, FastAPI, REST APIs**
Cloud Platforms: **GCP (Vertex AI, BigQuery, Cloud Storage), AWS (SageMaker, S3, EC2, EKS)**
MLOps & Workflow Orchestration: **MLflow, DVC, TFX, Kubeflow, Airflow**
DevOps & CI/CD: **Docker, Kubernetes, Jenkins, GitHub Actions, Terraform**
Model Training & Optimization: **Data Preprocessing, Feature Engineering, GPU Acceleration (CUDA), Automated Retraining, CI/CD**
Collaboration & Methodologies: **Agile/Scrum, Cross-functional teamwork**

Education and Certification

MSc. Computer and Information Science **Northumbria University** (London) ***2021-2023***

Supervised Machine Learning Stanford Online | **Python Programming Boot Camp**

Professional Experience

MLOps Engineer **Company-name** ***2023 – Present***

- **Designed, built, and maintained end-to-end ML pipelines** on **GCP (Vertex AI)** and **AWS**, integrating TensorFlow models and automating workflows, which improved model deployment speed by **~35%.**
- **Implemented real-time model monitoring and drift detection**, with automated alerts and retraining triggers that reduced downtime incidents by **~20%** and improved model reliability in production.
- **Optimized training and inference workflows using GPU acceleration (CUDA)**, reducing training time by up to **40%** and cutting compute costs by **~25%.**
- **Automated CI/CD workflows, including continuous training, testing, and redeployment of ML models**, with Jenkins, GitHub Actions, and Docker/Kubernetes, increasing deployment frequency by **~30%** while lowering release errors by **~15%.**
- **Applied TensorFlow Extended (TFX)** alongside **Kubeflow and Airflow** to orchestrate preprocessing, model validation, retraining, and deployment workflows, improving reproducibility and compliance.
- **Collaborated with cross-functional teams** (data scientists, engineers, product) to optimize data preprocessing and feature engineering, reducing data preparation time from hours to minutes and improving model accuracy by **~10%.**
- **Ensured high availability and scalability** of ML systems by applying MLOps best practices and automating infrastructure with **Terraform**, achieving **99.9% uptime** and seamless scaling for production workloads.