

# Nikhila Madhunala

nikhilamadhunala01@gmail.com | 304-460-5482 | linkedin.com/in/nikhila-madhunala

## PROFESSIONAL SUMMARY

---

AI/ML Platform Engineer skilled in Python, C++, and Go with expertise in MLOps tools like Kubeflow, MLflow, and SageMaker across AWS, Azure, and GCP. Builds scalable infrastructure and pipelines that enable rapid experimentation, reliable deployments, and measurable business value.

## EDUCATION

---

### Master's in Business Analytics

University of North Texas, Denton, TX – December 2024

### Bachelor of Technology in Computer Science

CMR College of Engineering and Technology, India – August 2021

## SKILLS & TECHNOLOGIES

---

**Programming:** Python, R, Java, C++, C#, Go, FastAPI, Kubernetes, Git, PyTorch

**Data Engineering:** Distributed systems, model training, model deployment, model monitoring

**Databases:** Redshift, BigQuery, Snowflake (extended with Amazon Redshift, Snowflake, and Google BigQuery warehouses)

**Cloud:** Amazon Web Services (AWS S3, AWS Lambda, AWS EC2, Amazon SageMaker), Microsoft Azure (Synapse, Azure Functions), Google Cloud Platform (BigQuery, Cloud Run)

**DevOps:** Kubeflow, MLflow, MLOps, CI/CD pipelines, containerization (Docker, Kubernetes), version control

**Analytics/ML:** Machine learning, deep learning, reinforcement learning, natural language processing, computer vision, classification, experimentation (A/B testing)

**Practices:** Agile methodologies, communication, collaboration, scalability, reliability, observability, optimization, prototyping, research, problem solving

## PROFESSIONAL EXPERIENCE

---

### AI ML Platform Engineer Intern

*Next Era Path*

*Feb 2025 – Present*

- Engineered a scalable ML platform on SageMaker and Kubernetes. Reduced deployment time by 40%, enabling seamless rollbacks and blue-green deployments, improving agility of model lifecycle management in production workloads.
- Built unified feature stores and registries to streamline workflows. Reduced duplication of effort and improved cross-team collaboration, enabling consistent feature use across ML development and inference environments.
- Developed cross-platform APIs with FastAPI for real-time prediction serving. Delivered sub-50ms latency and improved user experience while powering personalized recommendations for large-scale applications in production.
- Led cross-functional agile ceremonies and mentored junior engineers on MLOps. Enhanced team efficiency, improved collaboration, and raised organizational maturity in adopting MLOps practices across business-critical projects.

### AI ML Platform Engineer

*Infosys (Client: Westpac)*

*Jan 2021 – Dec 2022*

- Designed distributed training pipelines with PyTorch and Horovod across EC2 and Azure VMs. Cut training time by 30%, enabling efficient scaling and supporting training of larger, more complex ML models at scale.
- Automated model monitoring with MLflow and Prometheus. Reduced downtime by 25% through proactive detection and alerting, increasing system reliability and improving observability for production ML services.

- Collaborated with research scientists to design A/B experiments for evaluating ML models. Informed data-driven product decisions that increased customer engagement by 15% and improved adoption of new features.
- Documented monitoring pipelines, experiment results, and training methods for reproducibility. Shared across teams, improving transparency and accelerating iteration cycles for research and production ML workflows.

### **AI ML Platform Engineer Intern**

*Powersoft Global Pvt Ltd,*

*May 2020 – Aug 2021*

- Optimized infrastructure with right-sized resource allocation and adoption of spot instances. Delivered 20% annual cost savings while maintaining system reliability and scalability for production workloads.
- Partnered with platform and security engineers to align cloud configurations with compliance and governance. Ensured ML pipelines met enterprise standards while maintaining agility and operational performance.
- Designed CI/CD pipelines for multi-cloud ML deployments on AWS, Azure, and GCP. Improved release velocity, reduced manual errors, and standardized deployment workflows across hybrid environments for ML services.
- Conducted workshops on cloud cost management, container orchestration, and scaling strategies. Raised organizational awareness, improving cost efficiency and reliability of deployed ML systems in diverse workloads.

## **PROJECTS**

---

### **Capstone: Predicting Vehicle Recalls from Complaint Data**

*Aug 2024 – Dec 2024*

Inherited fragmented data pipelines across AWS, Azure, and GCP that caused duplication of effort. Consolidated infrastructure into unified pipelines for training and deployment. Delivered a single platform that improved scalability and reduced maintenance costs.

Implemented CI/CD pipelines integrating Kubeflow and MLflow for continuous training and deployment cycles.

Automated testing, validation, and release approvals. Significantly shortened model release cycles and increased business agility.

Drove adoption of observability practices by integrating monitoring with Prometheus and Grafana. Enabled teams to proactively track model drift and infrastructure health. Reduced time-to-detect and resolve issues, enhancing reliability.

Acted as a mentor for engineers transitioning into MLOps roles. Provided hands-on guidance in containerization, distributed computing, and cloud cost optimization. Built a knowledge-sharing culture that elevated overall technical maturity.