Nikhila Madhunala

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

Professional Summary

Engineer specializing in NLP and experimental ML systems with expertise in Python, FastAPI, and cloud services. Skilled in building classification models, scalable APIs, and evaluation frameworks. Focused on rapid experimentation, measurable outcomes, and reliable production deployment.

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

Technical Skills

Programming: Python, R, FastAPI, Docker, Kubernetes, Git

Data Engineering: Model deployment, model monitoring

Databases: Amazon S3, Amazon Redshift

Cloud: Amazon Web Services (Lambda, S3, EC2), Google Cloud Platform, Microsoft Azure

DevOps: MLflow, MLOps

Analytics/ML: Natural language processing, classification, experimentation, reinforcement learning

Practices: Communication, observability, experimentation, research

Professional Experience

NLP & Experimentation Engineer Intern
Next Era Path *Feb 2025 – Present*

- Built NLP classification models leveraging BERT architecture. Deployed models as serverless AWS Lambda functions, reducing infrastructure costs while maintaining consistently low latency for high-throughput production inference.
- Collaborated with linguists and domain experts to refine features. Enhanced model accuracy, demonstrated cross-functional teamwork, and delivered measurable improvements in applied NLP systems used in business applications.
- Designed clear annotation guidelines and documented best practices for NLP model training. Facilitated consistent labeling quality and improved reproducibility across projects by sharing knowledge with engineering and research peers.
- Conducted reinforcement learning research for content recommendation systems. Prototyped approaches that influenced roadmap planning and demonstrated the future potential of adaptive personalization for platform growth.

NLP & Experimentation Engineer
Infosys (Client: Westpac) *Jan 2021 – Dec 2022*

- Designed platforms for A/B testing and statistical analysis in ML workflows. Automated experiment processes to enable data-driven decision-making, reducing analysis time by 50% while ensuring rigorous statistical validity.
- Implemented real-time dashboards using Kibana and Grafana to monitor NLP model performance. Captured precision, recall, and latency metrics, supporting rapid detection of issues and maintaining reliable production services.
- Partnered with cross-functional teams to evaluate experimental outcomes. Translated insights into actionable improvements, ensuring ML models aligned closely with user experience and business value objectives.

- Defined best practices for ML experimentation workflows. Documented standards, improved adoption of consistent methodologies, and accelerated reliable experimentation cycles across the data science organization.

NLP & Experimentation Engineer Intern
Powersoft Global Pvt Ltd *May 2020 – Aug 2021*

- Created APIs using FastAPI for serving ML predictions. Integrated APIs into web services to improve user experience, increase adoption, and enable smooth interaction with NLP-powered features across applications.
- Assisted in migration from monolithic deployments to Kubernetes-based microservices. Improved scalability, resilience, and maintainability of ML systems, enabling faster iteration and streamlined operations in production.
- Automated CI/CD workflows to manage API releases and infrastructure rollouts. Reduced deployment errors, improved reproducibility, and ensured consistency across environments by embedding DevOps practices within ML workflows.
- Led internal workshops on API design, containerization, and observability practices. Shared expertise, raised team skill levels, and encouraged adoption of robust, scalable ML deployment patterns throughout engineering teams.

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.