SAHITH AITHA

Irvine, CA, 92614 |+1 (513)328-2956 |aithasahith0214@gmail.com|Linkedin |GitHub

PROFESSIONAL SUMMARY

MLOps Engineer with 3+ years of experience advancing intelligent, automated ML ecosystems across cloud platforms. Specialized in building reliable, explainable AI systems through Reinforcement Learning, survival analysis, and continuous evaluation frameworks. Experienced in unifying data quality, observability, and scalability to bridge research prototypes into production. Delivered impactful solutions including an ICU readmission prediction on 500K+ patient records. Strong academic foundation in AI and systems engineering from the University of Cincinnati, focused on delivering interpretable and efficient real-world ML solutions.

SKILLS

- **Machine Learning & AI:** Predictive Modeling, Survival Analysis, NLP, Computer Vision, Model Evaluation
- **Deep Learning:** TensorFlow, PyTorch, CNN, RNN, LSTM, Generative AI (VAE, GPT-based Models)
- **MLOps:** Kubernetes, Airflow, MLflow, CI/CD: GitHub Actions, Terraform (Certified Associate), Helm
- **Cloud Platforms:** AWS (DynamoDB, Lambda, EKS, SageMaker), GCP (Certified Professional Cloud Architect)
- **Data Engineering & Observability:** Python, SQL, OpenMetadata, REST APIs, NumPy, Prometheus, Grafana

PROFESSIONAL EXPERIENCE

MLOps Engineer, Preston Ventures LLC, Aliso Viejo, CA **June 2025 – Present**

- Built a model evaluation suite (AUC, Brier Score, Integrated Brier Score) for survival and risk models, improving interpretability and decision confidence across healthcare and asset-intelligence workloads.
- Implemented data quality monitoring across 30+ data assets using OpenMetadata and SQL-based validation checks, preventing data drift and ensuring consistent, high-quality training datasets.
- Automated the model lifecycle on AWS (EKS, S3, Lambda) orchestrated with Airflow, enabling continuous retraining, experiment tracking, and scalable, reliable production rollouts.

MLOps Engineer, University of Cincinnati, Cincinnati, OH **May 2024 – May 2025**

- Built and optimized ICU readmission prediction models (XGBoost, SVM, LightGBM) on the MIMIC-III dataset, achieving ~85% accuracy and enabling data-driven interventions for improved patient outcomes.
- Automated end-to-end ML pipelines with SageMaker, Git, and CI/CD, cutting model deployment cycle time by 30% and accelerating experimentation-to-production.
- Designed scalable ETL pipelines in PostgreSQL for healthcare records, reducing preprocessing overhead by 40% and ensuring faster, cleaner data availability for modeling.

Application Engineer, Amazon, India **January 2023 -June 2023**

- Built a real-time ticketing data processing application, to optimize internal processes by automating the detection and reporting of high-severity tickets, reducing manual effort by 15%.
- Streamlined ticket management processes by implementing workflows for JSON data ingestion and storage in Amazon DynamoDB, enhancing team efficiency and collaborating with cross-functional teams.
- Automated the process of parsing data and sending detailed reports via email, enabling managers to quickly assess team performance without reviewing each ticket manually.

EDUCATION

Master of Science in Information Technology **January 2024 - May 2025**

University Of Cincinnati, Cincinnati, OH

Course work: Artificial Intelligence and Machine Learning, Linux System Administration

PROJECTS

IntelliScale: RL-Driven Cloud-Native Auto-Scaling Platform **April 2025 – Present**

- Designed a containerized auto-scaling platform integrating Reinforcement Learning with Kubernetes. Used Prometheus metrics and REST APIs to optimize app responsiveness by 30% and cut over-provisioning by 25%. Deployed on GCP with vectorized state-action history for dynamic scaling decisions.

CPASS – Cincinnati Public Accessibility Systems-Survey **March 2025 – April 2025**

- Built a public data collection and visualization system using Flask + Roboflow for real-time urban analysis. Enabled secure multi-user survey access and real-time dashboard, increasing survey effectiveness by 30%.