

# Mayank Vyas

1265 E University Dr, Tempe AZ 85288

+1-602-517-1664 [mvyas7@asu.edu](mailto:mvyas7@asu.edu) [www.linkedin.com/in/mayank-vyas-369796213](https://www.linkedin.com/in/mayank-vyas-369796213) [github.com/Mayank-glitch-cpu](https://github.com/Mayank-glitch-cpu) <https://tinyurl.com/mayank-portfolio>

## EDUCATION

### Arizona State University

#### Masters in Data Science

Tempe, AZ  
Aug 2024 - May 2026

Statistical Learning, Natural Language Processing, Statistics, Cloud Computing, Big Data Analytics, Data Visualization

### Institute of Infrastructure Technology Research and Management

#### Bachelors in Electrical Engineering with minors in Computer Science

Ahmedabad, India  
Nov 2020 - May 2024

**Thesis:** MaskRoot: Deep Learning Pipeline for Root Phenotyping

**Relevant Courses** Data Structures and Algorithms, Computer Architecture, Artificial Intelligence, Distributed Systems

## SKILLS

**Programming Languages:** Python, Java, C++, SQL, JavaScript, Bash, HTML/CSS

**Cloud & DevOps:** Azure, Google Cloud, AWS, Docker, Kubernetes, Terraform, Jenkins, Ansible, Snowflake, Containerization, Data Warehouse, Cloud Infrastructure, DevOps, Assembly, backend, frontend, Api, code review

**Core Competencies:** Data Structures, Distributed Systems, Database Management, Operating Systems, CI/CD, System Design

**MLOps :** Tensorflow, PyTorch, Keras, CUDA, JAX, Golang

## PROFESSIONAL EXPERIENCE

### Software Engineer / Open Source Contributor

January 2025 – Present

#### Intel Automated Checkout System

- Added system performance feature by leveraging containerization (Docker) and orchestration, ensuring consistent deployments across cloud platforms (AWS, Google Cloud) while integrating centralized data warehouse solutions.
- Developed custom automation scripts and build tools using Jenkins and Ansible to publish the system's Real time data to grafana dashboards using MQTT Datasources
- Implemented robust logging, monitoring, and distributed system strategies, which improved overall system resilience and optimized cloud infrastructure for real-time operations.
- Spearheaded initiatives to upgrade legacy architectures by incorporating modern data structures, containerization, and DevOps best practices, visualize data to see system architecture performance at a glance.

### Machine Learning Assistant

May 2022 – January 2024

Indian Institute of Information Technology, Chennai (Sponsored by IIT Bombay)

- Developed an energy-efficient IoT and ML framework** leveraging LoRaWAN and Kalman filtering, reducing sensor energy consumption by 40% and bandwidth usage by 85%, enabling real-time decision-making in smart agriculture.
- Designed and deployed ML models on edge devices** using TensorFlow Lite and PyTorch, achieving 99.97% prediction accuracy while implementing model compression techniques that lowered power consumption by 82.89%.
- Optimized large-scale data aggregation and cloud integration**, reducing redundant data transmissions by 93.6% and cutting cloud transmission costs by 38%, while ensuring seamless deployment through automated CI/CD pipelines on AWS and Azure.
- Published innovative research on IoT-ML convergence**, with an APAE-based predictive framework tested on a 20-acre testbed, demonstrating a 35% faster response time and 82.89% energy savings at a tolerance threshold ( $\epsilon=1.0$ ).

## PUBLICATIONS

- DASA: An efficient [data aggregation algorithm](#) for LoRa enabled fog layer in smart agriculture. [Springer](#)
- On Reducing [Data Transmissions](#) in Fog Enabled LoRa Based Smart Agriculture. [IEEE](#)
- Intelligent [Data Forwarding Scheme](#) for LoRa based Fog Enabled Smart Agriculture. [IEEE](#)

## PROJECTS / OPEN-SOURCE

### Document Classification | Data mining, NLP, Machine Learning , GridSearch

February 2025

- Built a multi-class document classification system using NLP techniques (TF-IDF, BERT) and ML models (Decision Trees, Random Forest, AdaBoost), which achieved 92% accuracy on test data.
- Engineered a streamlined text pipeline by applying tokenization, stopwords removal, and lemmatization with NLTK, SpaCy, and Pandas, reducing the model error rate by 15%.
- Optimized model performance through hyperparameter tuning and cross-validation with GridSearchCV and PyTorch Lightning, boosting the F1-score from 0.82 to 0.89.
- Deployed the solution using Docker on Azure and Google Cloud and integrated it with a Snowflake SQL data warehouse, ensuring scalable and resilient data management.

### MASK ROOT: Deep Learning Pipeline for Root Phenotyping | Convolutional Neural Networks, OpenCV

May 2023 – January 2024

- Engineered an automated Mask R-CNN pipeline (TensorFlow/Keras) to segment primary roots and predict lengths in wheat, Arabidopsis, and rapeseed seedlings, achieving 96.5% segmentation accuracy with less than 10% variability.
- Pioneered transfer learning by deploying the pre-trained model on non-annotated Arabidopsis root datasets without retraining, achieving 95.2% mIOU and enabling scalable root analysis in resource-constrained settings.
- Integrated Otsu thresholding, Canny edge detection, and SAM for enhanced segmentation, attaining 99.8% detection accuracy.