Vivek Marri

+91 9390704655 | vivekyadavmarri@gmail.com | linkedin.com/vivek | github.com/vivek | Kothagudem, Telangana, India

Professional Summary

Aspiring AI/ML Engineer with a solid foundation in AI, Machine Learning, and MLOps. Skilled in developing and deploying intelligent applications using Python, FastAPI, Docker, and MLflow. Certified in AI and Generative AI (Oracle, NVIDIA). Passionate about creating scalable, real-world AI solutions.

Education

JNTUH University College of Engineering, Manthani Nov 2022 – May 2026(Expected)
Bachelor of Technology in Computer Science and Engineering

Experience

AI Intern — Applied Artificial Intelligence Track, TechSaksham Jan 2025 – Mar 2025
Microsoft & SAP (AICTE Joint CSR Initiative) *Virtual*

- Developed and deployed an **NLP-based Healthcare Chatbot** using **Python**, **FastAPI**, and **Docker**, with modular training and inference pipelines enabling production-style deployment.
- Implemented supervised ML models and neural network pipelines using Scikit-learn and TensorFlow, integrating preprocessing and feature engineering workflows. .
- Contributed to design reviews and iterative improvements aligned with the initiative's **AI-for-Social-Good** objectives.

Projects

Urban Flood Early Warning System — MLOps Platform for Real-Time Prediction **GitHub**

- Developed a real-time flood forecasting platform to provide early warnings for urban areas, improving public safety and emergency response times.
- Designed a real-time flood forecasting pipeline with automated **MLOps CI/CD**, using **Apache Airflow** for scheduled retraining, drift detection, data validation, and model versioning via **MLflow**.
- Deployed a high-throughput, low-latency inference service with **FastAPI + Docker Compose**, achieving **sub-50ms response times** and supporting horizontal scaling in production.
- Engineered an **ACID**-compliant **PostgreSQL** backend for prediction logging, traceability, and automated monitoring hooks, enhancing reliability and observability.
- Optimized inference performance by resolving **NumPy** serialization bottlenecks, improving peak throughput by **3×** under stress tests.

Intelligent Logistics & Delivery Optimization (IntelliLog-AI) **GitHub**

- Developed a solution to optimize delivery operations, reduce delays, and balance driver workloads for urban logistics fleets.
- Built a hybrid **ML + optimization** pipeline using **XGBoost Regression** for delivery-time prediction and a custom **VRP solver** for multi-driver route optimization.
- Implemented **graph-theoretic heuristics** (Dijkstra, A*) and **Google OR-Tools** with capacity/time-window constraints, reducing delivery delays by up to **25%** in simulated environments while balancing workloads.
- Deployed a low-latency microservice with **FastAPI** and **Docker**, supporting real-time route generation and ETA predictions with <300ms p95 latency.
- Built an interactive **Streamlit Dashboard** for real-time route visualization, fleet tracking, operational KPIs, and **SHAP**-based model interpretability.

Technical Skills

ML/AI Frameworks: TensorFlow, PyTorch, Scikit-learn
MLOps & Deployment: Apache Airflow, Docker, MLflow Registry, FastAPI, Streamlit, Git/GitHub
Languages & Core: Python, C, SQL, JSON, YAML
Databases & Data Tools: PostgreSQL, MongoDB, MySQL, Pandas, NumPy

Certifications, Achievements & Leadership

- Earned the **Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate** credential from **Oracle University**, validating proficiency in **AI**, **ML**, and **OCI** — View Credential.
- Earned the **Certificate of Competency in Generative AI with Diffusion Models** by **NVIDIA**, demonstrating applied expertise in **Diffusion Techniques** and **Generative AI Frameworks** — View Credential.
- **Technical Fest Head & Tech Lead** for the Annual Technical Fest at JNTUH UCEM (2025).
- **Organizing Committee Head** for the National Level Workshop on Machine Learning and Big Data at JNTUH UCEM (2024).