Venkata Sai Saravan Pathapati

410-869-5848 saravan.pathapati@gmail.com linkedin.com/in/saravan-pathapati saravan-portfolio

Professional Summary

AI/ML Engineer with hands-on experience in developing scalable AI solutions using deep learning frameworks like PyTorch and TensorFlow. Skilled in machine learning techniques including NLP, computer vision, generative models, and MLOps for deploying real-time applications in cloud environments. Proven track record of collaborating on cross-functional teams to deliver efficient, secure systems, optimizing performance with quantifiable improvements in accuracy and latency.

Technical Skills

Languages: Python, JavaScript, C#, Bash
Frameworks & Libraries: PyTorch, TensorFlow, OpenAI API, XGBoost, Scikit-learn, FastAPI, NumPy, Pandas, Matplotlib, Dask
ML Techniques: Transformers (BERT, T5), CNNs (ResNet, AlexNet, MobileNet), GANs, Diffusion Models, RNNs, LSTMs, GNNs, Q-Learning, DQN, BYOL, CLIP, SVM, PCA, K-Means, Transfer Learning
MLOps & Cloud: GCP (Vertex AI, Pub/Sub, BigQuery), AWS (S3, Lambda, SageMaker), Docker, Kubernetes, CI/CD, REST APIs
Data & Other: SQL, Feature Engineering, A/B Testing, EDA, Hyperparameter Tuning, TDD, SMOTE, KL Divergence, SHAP, LIME, GitHub Actions

Experience

OQP Solutions **DEC 2024 – Present**
AI Engineer *Washington DC, USA*

- Built a comprehensive NLP pipeline utilizing BERT, Transformers, and OpenAI for effective real-time threat detection on GCP.
- Deployed CNN and RNN models for predictive analytics with FastAPI, Docker, and Kubernetes using robust CI/CD workflows.
- Created an generative chatbot employing GANs and Diffusion models integrated with CLIP for zero-shot, multi-modal reasoning.
- Reduced system latency through GPU acceleration, applying GNNs and Q-Learning techniques for inference-time optimization.
- Improved computer vision accuracy by 15% via ResNet transfer learning and training with advanced data augmentation methods.

Technologies: Python, GCP, OpenAI, PyTorch, TensorFlow, FastAPI, SQL, Docker, Kubernetes, XGBoost

UNIIAN **JAN 2024 – MAY 2024**
Machine Learning Engineer *Texas, USA*

- Engineered secure FastAPI endpoints and preprocessing pipelines using Dask to deploy reliable ML pipelines efficiently.
- Built a robust data migration system incorporating DP and graph algorithms with CI/CD orchestration on Kubernetes platform.
- Optimized model validation processes using PCA, XGBoost, and SVMs for enhanced precision and improved runtime performance.
- Integrated NLP models including RNN/LSTM and RL with comprehensive OpenAPI documentation and test-driven development.
- Reduced data errors by 30% employing K-Means clustering and self-supervised learning for effective anomaly detection.

Technologies: Python, FastAPI, Dask, Docker, Kubernetes, SQL, PyTorch, TensorFlow, NumPy, Pandas

Dazarus Private Limited **Jul 2023 – Dec 2023**
Machine Learning Engineer Trainee *Remote, India*

- Built ML models for IoT using XGBoost and TensorFlow; deployed CNNs for anomaly detection and edge efficiency improvement.
- Performed exploratory data analysis and tuning for IoT sensor data using Pandas, NumPy, and Scikit-learn-based workflows.
- Enabled real-time GNN-based alerts and RL monitoring integrated with GCP Pub/Sub for intelligent smart IoT workflows.
- Implemented Dockerized MLOps and probabilistic models to enhance inference robustness under various uncertainty conditions.
- Applied Bayesian methods to quantify uncertainty and optimize edge model performance during deployment scenarios.

Technologies: Python, TensorFlow, XGBoost, Pandas, NumPy, Scikit-learn, GCP, Pub/Sub, Docker

Projects

LinkedIn SafeGuard **Dec 2023**
Git: https://github.com/saravan-p/LinkedIn_Safeguard

- Developed ML model to detect fraudulent profiles using Decision Trees and SMOTE, achieving 95.65% accuracy.
- Built resume scraper to mask sensitive data based on risk scores and context-aware redaction rules.

Education

- **University Of Maryland** Sep 2022 – Jun 2024
  Masters in Computer Science – GPA: 3.3, Baltimore County, MD
- **Vellore Institute Of Technology** Jun 2018 – May 2022
  Bachelor's of Technology in Computer Science – GPA: 8.42, Vellore, IN

Publications & Certifications

- **Artificial Intelligence in Game Programming** — *ICT4SD 2022, Springer*, Nov 2022
  DOI: https://doi.org/10.1007/978-981-19-5221-0_60
- **Google Cloud Professional Machine Learning Engineer**
  Issued: Oct 2024 | Expires: Oct 2026
  https://www.credly.com/badges/987f082e-ef30-4b72-bde3-9cd860c3dd08