Lokesh Kanna Rajaram

Email | Portfolio | LinkedIn | LeetCode |+1(716) 520 7064| GitHub

Summary

Data scientist with expertise in machine learning, deep learning, and NLP using Python, PyTorch, and TensorFlow. Experienced in developing predictive models, generative AI systems, and computer vision pipelines for real-world applications. Combines statistical rigor and engineering skills to deliver interpretable and high-impact AI solutions.

TECHNICAL SKILLS

Programming Languages & Databases: Python, R studio, R, MySQL, Pytorch, Hadoop, Apache Spark, Kafka, Pandas, Natural Language Processing, Predictive Modeling, MongoDB, Oracle, PostgreSQL, Machine Learning/AI.
Tools & Platforms: Docker, GitHub, PowerBI (DAX), Tableau, Microsoft Office, Azure, Generative AI, AWS,S3,ETL Pipelines, CI/CD Pipelines, LLMs, Statistical Analysis, Cloud Infrastructure, Big Data, Data Visualization.
AWS Certifications: Cloud Practitioner – CLF-CO2, Cloud-based AI Solutions

EDUCATION

University at Buffalo, The State University of New York NY, USA **Dec 2025**
Master of Science, Data Science
Coursework: Numerical Mathematics, Statistical Data Mining, Database Fundamentals, Data Intensive Computing, Data Model Query Languages, Introduction to Machine Learning, Computer Vision.

WORK EXPERIENCE

Data Scientist Intern – Nissha Medical Technologies ,Buffalo, New York **Aug 2025 – Dec 2025**
- Developed a Convolutional Neural Network (CNN) in PyTorch with OpenCV image processing, reducing the Q-Block defect rate by 25% (from 4% to 3%) on a 30 million ticket-per-day production line.
- Conducted exploratory data analysis (EDA) using Pandas and Matplotlib to correlate production metadata with image defects, Identifying key variables such as machine condition and environmental factors that informed data augmentation strategies and improved model performance.
- Deployed a containerized vision model using Docker in a pilot test, reducing machine downtime by 20% and validating the model's impact, aligning with business optimization goals and setting a foundation for broader implementation.

PROJECTS

Generative Models Benchmarking GenAI Models: Gans, Vaes, And Diffusion Models **Aug 2025**
- Developed and fine-tuned a U-Net-based DDPM in PyTorch, achieving competitive FID and IS scores on MNIST, CIFAR-10, and CelebA compared to VAE and DCGAN baselines.
- Led the evaluation and analysis phase, benchmarking convergence, stability, and sample diversity across models using FID, Inception Score, and qualitative visual comparisons.
- Streamlined training and visualization pipelines, reducing experiment setup time by 30% through modular scripts, automated checkpoints, and comparison notebooks.

Real-Time Traffic Vehicle Detection And Counting System **Jul 2025**
- Developed a Flask-based web application integrating YOLOv5, OpenCV leveraging deep learning for real-time object detection and classification achieving with vehicle classification accuracy of 83%.
- Engineered a scalable backend pipeline to ingest and transform video streams, enabling live frame annotation, vehicle detection, and count aggregation using machine learning algorithms.
- Designed and optimized a frame-sampling and streaming pipeline to process high-volume video data (every 5th/10th frame), reducing computational load while preserving analytical accuracy.

Amazon Book Review Using Big Data Pipeline **May 2025**
- Designed and implemented a scalable big data ETL pipeline development using Hadoop, Pyspark, and Docker, enabling data ingestion and ML-ready transformation of over 1 Millon records.
- Automated ingestion of CSV data into HDFS with CLI tools, ensuring fault-tolerant parallel access across nodes.
- Accelerated model training by 40% using optimized text processing (Tokenizer, StopWordsRemover, HashingTF, IDF) in Spark ML, achieving up to 90.4% accuracy.

Optimized Bulk Stock Selling Strategies With Machine Learning **Dec 2024**
- Boosted stock price and volume prediction accuracy by 15% (to 78%) deploying end-to-end machine learning pipelines with Random Forest, Gradient Boosting, LSTM, and regression models on a 4-year NVIDIA dataset.
- Implemented algorithmic trading strategies (VWAP, TWAP) to reduce market impact during bulk selloffs by applying statistical methods and data modelling and real-time data engineering techniques.
- Enriched model interpretability with technical indicators (RSI, Bollinger Bands), conducting regression analysis, and clustering through time series analysis and data visualizations.