

# Lokesh Kanna Rajaram

[lrjaram@buffalo.edu](mailto:lrjaram@buffalo.edu) | [Portfolio](#) | [LinkedIn](#) | [LeetCode](#) | +1(716) 520 7064 | [GitHub](#)

Data Science expert and certified Cloud Practitioner skilled in building scalable data solutions, predictive models, and cloud pipelines that drive actionable business insights. Proficient in machine learning, big data, and visualization tools to solve complex problems and power data-driven decisions.

## TECHNICAL SKILLS

---

**Programming Languages & Databases:** Python, R shiny, MySQL, Pytorch, Hadoop, Apache Spark, Map Reduce, Pandas

**Tools & Platforms:** Docker, GitHub, Git, PowerBI, Tableau, Microsoft Excel, Generative AI, AWS

**AWS Certifications:** Cloud Practitioner – CLF-CO2

## EDUCATION

---

University at Buffalo, The State University of New York

NY, USA

*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 (Local agricultural initiative) – Volunteer, India

05/2025 – 08/2025

- Increased crop yield by 18% and reduced fertilizer use by 22% through EDA, K-Means clustering, and predictive modeling (Random Forest, XGBoost) on historical data.
- Automated irrigation scheduling with a regression model, cutting labor costs by 15%, and delivered insights via PowerBI dashboards, seaborn, and matplotlib visualizations.
- Developed and optimized complex SQL queries with joins, aggregations, and window functions to integrate soil, weather data, and yield data. Collaborated with agronomists to align data extraction with domain-specific insights.

## PROJECTS

---

Generative Models Benchmarking GenAI Models: GANs, VAEs, and Diffusion Models ([link](#))

08/2025

- 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 ([link](#))

07/2025

- 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 ([link](#))

05/2025

- Designed and implemented a scalable big data ETL pipeline development using Hadoop, Pyspark, and Docker, enabling data ingestion and ML-ready transformation of over 1Million 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 ([link](#))

12/2024

- 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.