

Lokesh Kanna Rajaram

lrjaram@buffalo.edu | [Portfolio](#) | [LinkedIn: Lokesh kanna](#) | [LeetCode](#)

+1(716) 520 7064 | [GitHub](#)

EDUCATION

University at Buffalo, The State University of New York

Buffalo, NY, USA

Master of Science

- **Course:** Engineering 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

05/2025 – Present

- Increased crop yield by 18% by performing data wrangling and statistical analysis on historical weather, soil, and yield data, and developing predictive models using Random Forest and XGBoost.
- Reduced fertilizer consumption by 22% through unsupervised learning techniques (K-Means clustering) on soil nutrient profiles to inform precision fertilization strategies.
- Minimized irrigation-related labor costs by 15% by building a regression-based model that automated watering schedules using real-time sensor data.

PROJECTS

REAL-TIME TRAFFIC VEHICLE DETECTION AND COUNTING SYSTEM

07/2025

- Developed a Flask-based web application for automated traffic analysis using YOLOv5, OpenCV leveraging deep learning for real-time object detection and classification achieving with vehicle classification accuracy of 83%.
- Engineered a scalable backend pipeline for video ingestion, transformation, and live frame streaming, integrating machine learning logic for vehicle categorization and efficient count aggregation with real-time annotation.
- 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

05/2025

- Designed and implemented a scalable big data pipeline to process Amazon book reviews using Hadoop and Apache Spark on Docker clusters, enabling distributed ingestion, and ML-ready transformation of over 1 million records.
- Automated data ingestion and storage into HDFS from raw CSV sources using CLI tools and batch upload strategies, ensuring robust fault-tolerant storage for parallel data access across nodes.
- Optimized text feature pipelines with Tokenizer, StopWordsRemover, HashingTF, and IDF in Spark ML library, balancing dimensionality, which accelerated downstream model training by 40% and achieved up to 90.4% accuracy.

OPTIMIZED BULK STOCK SELLING STRATEGIES WITH MACHINE LEARNING

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.
- Built efficient and scalable algorithmic trading strategies (VWAP and TWAP) to reduce market impact and enhance profitability for bulk stock transactions, applying statistical modeling and real-time data engineering techniques.
- Enhanced feature quality and model interpretability by integrating advanced technical indicators (RSI, Bollinger Bands), conducting regression analysis, and clustering through time series analysis and visualizations.

TEXT SUMMARIZER USING DEEP LEARNING

07/2024

- Improved system scalability and performance for text summarization models by deploying secure, high-availability infrastructure on AWS EC2 using IAM, and optimizing performance for complex datasets.
- Accelerated deployment cycles and reduced manual overhead by implementing an automated CI/CD pipeline with GitHub Actions and containerizing machine learning models using Docker for reproducible builds.
- Delivered a production-ready, user-centric summarization tool by integrating advanced machine learning algorithms, managing cloud-based deployment on EC2, automating workflows, and collaborating cross-functionally for seamless continuous delivery.

TECHNICAL SKILLS

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

Tools & Platforms: AWS, Docker, GitHub, Git, GCP, PowerBI, Tableau, Ms-Excel