

RAGUL NARAYANAN MAGESH

Boston, MA (Open to Relocation) | magesh.ra@northeastern.edu | [Portfolio](#) | [LinkedIn](#) | [GitHub](#) | +1 (617) 792-7068

EDUCATION

Northeastern University, Boston, MA

Sep 2024 - May 2026

Master of Science in Data Analytics Engineering, GPA - 3.8 / 4.0

Relevant Coursework: Neural Networks & Deep Learning , NLP, Machine Learning Operations, Data Mining, Generative AI

TECHNICAL SKILLS

Languages & Frameworks: Python, SQL, PySpark, TensorFlow, PyTorch, Keras, Scikit-learn, Hugging Face Transformers

Generative AI & LLMs: FLAN-T5, OpenAI API, LangChain, Prompt Engineering (zero/few-shot), RAG, Fine-tuning

MLOps: MLflow, DVC, Airflow, Docker, CI/CD, GCP (BigQuery, Vertex AI), AWS (SageMaker, S3), Experiment Tracking

NLP: BERT, RoBERTa, BM25, Vector Embeddings, Cross-Encoder Reranking, Sentiment Analysis

Data Engineering: Snowflake, PostgreSQL, MongoDB, Star Schema, ETL/ELT Pipelines | **Visualization:** Tableau, Power BI

PROJECTS

[EchoAI - Scalable LLM-Powered Review Response System](#) [GCP, Airflow, BigQuery, MLflow, Streamlit, FLAN-T5] Dec 2025

- Architected and deployed an **end-to-end MLOps pipeline** processing 100K+ customer reviews using FLAN-T5 LLM for generative AI-powered automated response generation with **85%+ confidence scores**, integrated sentiment analysis, bias detection, and anomaly detection modules
- **Built production-grade ML infrastructure on GCP** with Airflow orchestration, BigQuery data warehousing, MLflow experiment tracking, DVC versioning, and CI/CD-driven deployment to a Streamlit inference application

[Amazon Electronics Product Recommendation System](#) [PyTorch, NLP, AWS SageMaker, FastAPI, Qdrant, GCP] Dec 2025

- Engineered a **hybrid recommendation system** processing 6.7M+ reviews across 31,100 products using BM25 sparse retrieval combined with BGE dense vector embeddings and cross-encoder reranking, achieving NDCG@10 of 0.854
- Deployed **Pegasus abstractive summarization** and **RoBERTa sentiment analysis** on AWS SageMaker; served end-to-end pipeline via FastAPI with Qdrant vector store for scalable semantic search

[Convolutional Neural Network for Multi-Class Object Classification](#) [TensorFlow, Keras, Computer Vision, CNN] Jan 2026

- Designed and evaluated **20 CNN architectures for 39-class image classification**, systematically optimizing dropout regularization, batch normalization, learning rate schedulers, and optimizer configurations to **achieve 96.59% validation accuracy**
- Applied rigorous experiment design methodology to quantify performance trade-offs across architectures, translating findings into a reproducible model selection framework

[Custom Object Detection using YOLOv8 OBB](#) [YOLOv8, Roboflow, PyTorch, Annotation, Data Augmentation] Jan 2026

- Built a **YOLOv8 oriented bounding box (OBB) detection model** for 39-class object detection on 3,900+ custom-annotated images using Roboflow, applying multi-stage data augmentation and training on **NVIDIA A100 GPU**
- Validated real-world multi-object detection performance on custom collage test images, demonstrating robust generalization across complex, densely packed scenes

WORK EXPERIENCE

Data Scientist

Aug 2023 - Jul 2024

360DigitMG

Chennai, India

- Analyzed IQVIA pharmaceutical sales data to benchmark product performance across therapeutic categories, applied predictive analytics and HCP profiling to forecast sales and identify high-value prescribers, increasing **targeting precision by 25%**
- Modeled and transformed large-scale datasets in Snowflake using advanced SQL, **improving reporting efficiency by 40%** and enabling seamless integration with Tableau dashboards for real-time decision-making
- Developed interactive Tableau dashboards visualizing prescription activity (NRx & TRx), regional sales trends, and KPIs, directly supporting strategic planning for senior stakeholders

Data Engineer

Apr 2022 - Jul 2023

Wipro

Bengaluru, India

- **Migrated 150+ Teradata objects** (retail, customer, supply-chain data) to Snowflake, ensuring metadata alignment and full data consistency; designed Star Schema models optimized for downstream analytics
- Built and optimized PySpark and Airflow ETL pipelines, **reducing processing time by 30%** and automating incremental data loads for analytics-ready delivery
- Developed Python-based semantic monitoring tools to validate data quality, enforce business rules, and trigger real-time alerts, ensuring system reliability across production pipelines