

Mano Harsha Sappa

Fairfax, VA (Open to Relocation) | <https://www.linkedin.com/in/manoharshasappa/> | 571-663-9406 | sappamanoharsha@gmail.com | [Portfolio](#)

PROJECTS

LLM Engineer - [\[GitHub\]](#)

Virginia, USA

DARIA-3o: Multi-Modal AI Chatbot for Tunnel NDE Queries

Feb 2025 – Apr 2025

- Built a domain-specific chatbot using LLaMA and Phi2 LLMs integrated with FAISS for semantic search, achieving 96% accuracy and 92% relevance in tunnel NDE query resolution.
- Engineered a structured QA dataset from InfoTunnel web content using BeautifulSoup, regex, and POS tagging, then fine-tuned LLMs using LoRA + PEFT for improved domain understanding.
- Deployed the solution using Streamlit (frontend), LangChain (context orchestration), Whisper (voice input), and Hugging Face Transformers for scalable, multimodal interaction.

Data Scientist - [\[GitHub\]](#)

Virginia, USA

Bridges at Risk: ML Analysis of Environmental Impact on Bridge Health

Nov 2024 – Jan 2025

- **Predicted bridge deterioration risks** by clustering 34K+ FHWA records with K-Means (Silhouette Score = 0.87) and classifying risk with XGBoost and MLP (77.7% accuracy), aiding public infrastructure strategy.
- **Extracted actionable maintenance insights** via PySpark-based EDA, PCA, and feature engineering, supporting regional policy prioritization across environmental and structural factors.

Machine Learning Engineer - [\[GitHub\]](#)

Virginia, USA

NYC Recycling Efficiency (2016–2019) – AWS-Based Pipeline

Aug 2024 – Oct 2024

- **Delivered a scalable data quality pipeline** by engineering capture-efficiency metrics, cleaning and standardizing 2800+ NYC recycling records using Python on AWS EC2, and exporting structured datasets to S3—enabling reliable downstream analysis and reporting.
- **Enabled zone-level policy insights** by ranking borough-wise KPIs via SQL and R, and visualizing Data Quality Scorecards, supporting city leadership with seasonally-aware recycling performance dashboards.
- **Strengthened audit-readiness and trend forecasting** by integrating AWS Glue DataBrew, MySQL, and R into a modular pipeline architecture for anomaly detection, data profiling, and ML deployment readiness.

Cloud Data Engineer - [\[GitHub\]](#)

Virginia, USA

Cloud-Based Predictive Analytics and IoT Stream Processing using Azure Databricks

May 2024 – Jul 2024

- **Developed a Spark SQL analytics pipeline** in **Azure Databricks** to uncover department-level attrition trends using grouped queries and visual dashboards, enabling HR teams to monitor workforce stability.
- **Built end-to-end cloud infrastructure** with **Azure Virtual Machines**, IoT Hub, and Stream Analytics to simulate real-time telemetry processing and ingest sensor data into Azure Storage for smart system analytics.
- **Improved retail strategy insights** by modeling product affinity with FP-Growth and predicting employee churn using PySpark Logistic Regression, isolating key drivers like OverTime and JobSatisfaction via MLlib pipelines.

Machine Learning Engineer - [\[GitHub\]](#)

Virginia, USA

SupplySight: AI-Based Warehouse Product Prediction

Jan 2024 – Apr 2024

- **Enabled intelligent warehouse planning** by developing a Linear Regression model ($R^2 = 99.2\%$) on 25K+ records to forecast warehouse output, streamlining supply chain decision-making.
- **Built real-time predictive infrastructure** using MCP + UV, deploying model tools with Python, Pandas, Pickle, and YAML, and delivering CLI and web interfaces for operational teams.

WORK EXPERIENCE

Data Science Associate - [\[GitHub\]](#)

Bengaluru, India

CORIZO Edutech Private Limited

May 2022 – Dec 2022

- **Enabled video performance forecasting** by building a Random Forest and SVR-based regression pipeline on 16K+ YouTube records, achieving up to 0.89 R^2 and supporting data-driven content decisions.
- **Boosted model reliability by 30%** through feature engineering, correlation filtering, and GridSearchCV-tuned hyperparameters, validating performance via R^2 , MSE, and importance plots.

EDUCATION

George Mason University

Fairfax, VA

Master of Science in Data Analytics Engineering (GPA: 3.93)

Expected Graduation: December 2025

Relevant Coursework: Machine Learning & AI, Data Engineering & Big Data, Analytics & Business Intelligence.

Skill:

Data & BI Tools: SQL, Excel, Power BI, Tableau, Data Visualization, Exploratory Data Analysis, Joins & Aggregations, ER Modeling, Data Normalization, MS SQL Server, MySQL, PostgreSQL

Machine Learning & AI Development: Python, NumPy, scikit-learn, Feature Engineering, Model Evaluation, Model Tuning, Hugging Face, OpenAI API, ChatGPT, Transformers, BERT, RAG Pipelines, LangChain, Prompt Engineering, Vertex AI

Tools & Platforms: Jupyter, Google Colab, VS Code, Adobe XD, Azure IoT Hub, Azure Data Factory, AWS EC2/S3/Glue