Mano Harsha Sappa

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

PROJECTS

Machine Learning Engineer - [GitHub] **Virginia, USA**

SupplySight: AI-Based Warehouse Product Prediction June 2025 – June 2022

- Developed a full-stack ML system predicting warehouse output (in tons) using Linear Regression (R^2 = 99.2%) on 25,000+ records, enabling smarter logistics and resource planning.
- Performed extensive preprocessing: IQR-based outlier capping, Box-Cox transformation, one-hot encoding of categorical features, and feature scaling to ensure robust model input.
- Deployed the model via a real-time MCP + UV interface, building CLI and web-based prediction tools using Python, Pandas, Pickle, and YAML-based server configuration.

LLM Engineer - [GitHub] **Virginia, USA**

DARIA-3o: Multi-Modal AI Chatbot for Tunnel NDE Queries Jan 2025 – May 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 Sep 2022 – Dec 2022

- Analyzed 34,000+ FHWA bridge records using PySpark, clustering them into four risk levels with K-Means (Silhouette Score = 0.87) based on environmental and structural features.
- Built and evaluated classification models (XGBoost, Random Forest, MLP) to predict bridge risk levels, achieving 77.7% accuracy and identifying key features like age, ADT, and precipitation.
- Applied advanced EDA, feature engineering, and PCA for dimensionality reduction, enabling clearer insights into regional deterioration patterns and maintenance prioritization.

Data Analyst - [GitHub] **Virginia, USA**

Traffic Collision Analysis: Montgomery County, MD Sep 2022 – Dec 2022

- Analyzed 172,000+ crash records to predict injury severity using Random Forest, identifying key risk factors like collision type, weather, and road surface with high model accuracy.
- Engineered features from timestamp and categorical variables, performed data cleaning in R (e.g., factor conversion, NA removal), and visualized trends using ggplot2.
- Validated model using confusion matrix, variable importance plots, and chi-squared tests, producing actionable insights for road safety and policy intervention.

WORK EXPERIENCE

Data Science Associate - [GitHub] **Bengaluru, India**

CORIZO Edutech Private Limited May 2022 – Dec 2022

- Built a predictive analytics pipeline using Random Forest and SVR, achieving up to 0.89 R^2 for like/view prediction on 16K+ YouTube videos, enabling data-driven content strategy.
- Engineered and cleaned features by removing low-informative fields, performing correlation analysis, scaling, and transforming time-based data for optimal regression input.
- Conducted EDA and hyperparameter tuning with GridSearchCV, improving SVR accuracy by 30% and validating models using R^2, MSE, and feature 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, Deep Learning, Natural Language Processing, Data Mining, Artificial Intelligence, Data Structures & Algorithms, Database Systems, Big Data Analytics, Cloud Computing, Statistics for Data Science, Computer Vision, Software Engineering, Data Visualization, Advanced Python Programming, Distributed Systems, Business Intelligence Systems

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, Model Evaluation, Model Tuning, Feature Engineering, Hugging Face, OpenAI API, ChatGPT, Transformers, BERT, RAG Pipelines, LangChain, Prompt Engineering, Vertex AI

Tools & Platforms: Jupyter, Google Colab, VS Code, Adobe XD