

SURYA VARDHAN REDDY PUCHALAPALLI

✉ surya.unf257@gmail.com
🌐 linkedin.com/in/suryavardhanreddypuchalapalli
🌐 surya1332002.github.io

📄 github.com/surya1332002
📍 Houston, TX, USA
📞 +1 346-628-8609

Profile

Master's student in Engineering Data Science with experience in Generative AI, LLMs, and end-to-end ML pipelines. Skilled in Python, SQL, Power BI, PyTorch, and LangChain. Passionate about solving real-world problems at scale through collaborative AI research and intelligent systems development.

Education

Masters in Engineering Data Science, University of Houston — GPA: 3.63/4.0 *Aug 2023 – May 2025*
B.E. in Computer Science, Sathyabama University — GPA: 8.81/10 *Jun 2019 – Apr 2023*

Experience

University of Houston IT – Data Infrastructure & Analytics (IA) *Jan 2024 – Present*

- Analyzed Wi-Fi usage data and campus survey feedback from 6,500+ access points and Cape sensors to drive strategic infrastructure upgrades across 60% of campus
- Built a Power BI automation pipeline that unified data from multiple sources, enabling auto-refresh dashboards; reduced manual reporting time by 80% for weekly team meetings
- Collaborated on network redesign and AP reconfiguration plans using usage insights; contributed to phased wireless expansion, targeting 80% campus coverage by 2025

Projects

Gen AI Cold Email Generator (LLM + LangChain + Semantic Search) *Mar 2025*

- Designed a generative email automation tool using Llama 3.3, LangChain, and ChromaDB
- Implemented semantic similarity scoring and vector-based retrieval using real-time scraping and embedding pipelines
- Applied prompt engineering for response control and relevance; achieved 93% semantic accuracy on evaluation set

Movie Recommendation System (KNN + Association Rule Mining) *Nov 2024*

- Engineered a hybrid recommendation engine using KNN and Apriori algorithm for collaborative filtering and association rule mining on the MovieLens 20M dataset; developed an interactive R Shiny web interface
- Reduced dataset to 3.8M rows through optimized preprocessing; conducted user similarity analysis and basket optimization to generate contextual suggestions
- Achieved a 71.4% top-5 hit rate and 22% increase in recommendation diversity over the KNN baseline; evaluated using precision@k and hit rate

Background Subtraction System for Real-Time Video Streams *May 2024*

- Developed a foreground detection model for video frames using multi-frame averaging and adaptive thresholding
- Applied morphological filters and OpenCV to enhance detection accuracy with latency under 120ms per frame
- Ideal for use in embedded/edge AI systems such as smart cameras

Binary Image Classification for Oil Spill Detection *Jul 2024*

- Built a CNN model (VGG-19) to detect oil spills in satellite ocean imagery
- Achieved 94% validation accuracy and optimized with data augmentation and dropout regularization
- Deployed model to classify high-resolution binary aerial frames for environmental monitoring

Research

Fish Species Classification using CNN Algorithm

Published and presented at ICCMC 2023, showcasing a deep CNN model for multi-class fish species recognition. Used structured labeling, data preprocessing, and VGG-inspired architecture for classification accuracy above 91%. Focused on domain adaptation challenges and optimization for real-time applications

Skills

Programming Tools: Python, SQL, Power BI, Git, VS Code, Bash

Libraries & Frameworks: PyTorch, TensorFlow, Scikit-learn, Keras, OpenCV, LangChain, ChromaDB

ML Techniques: Supervised Learning, Unsupervised Learning, Reinforcement Learning, Transfer Learning

Concepts: LLM Fine-tuning, Prompt Engineering, Feature Engineering, Semantic Search, Model Evaluation

Platforms: Jupyter, Google Colab, GitHub, AWS (familiar), OCI (certified), Tableau (familiar)

Visualization: Power BI, Matplotlib, Seaborn

Databases: Oracle, ChromaDB, MySQL

Research & Collaboration: Technical Writing, Cross-functional Teams, Data Storytelling