SHYAM HARI

Pathanamthitta, Kerala, India | +91 9400563384 | shyamhwrk@gmail.com
linkedin.com/in/shyam-hari-5389492b3 | github.com/shyamhari1074

Professional Summary

Data Scientist and ML Engineer pursuing B.Tech in AI & Data Science with hands-on experience building, evaluating, and deploying machine learning systems across supervised learning, unsupervised learning, deep learning, reinforcement learning, and LLM-powered pipelines. Proficient in Python, PyTorch, and Scikit-learn with strong foundations in data preprocessing, feature engineering, statistical modeling, and model evaluation. Proven ability to ship end-to-end ML systems to production — from raw data ingestion through model training to deployed, user-facing applications. Experienced integrating LLM APIs, building FastAPI backends, deploying on Vercel, and delivering actionable insights to stakeholders.

Skills

Programming	Python, SQL, JavaScript, Java, C
ML Frameworks	PyTorch, Scikit-learn, Stable-Baselines3, HuggingFace Transformers
Core ML	Supervised Learning (Regression, Classification), Unsupervised Learning (Clustering, K-Means), Reinforcement Learning (PPO), Neural Networks, Genetic Algorithms, Computer Vision
Data Science	Pandas, NumPy, EDA, Feature Engineering, Data Wrangling, Statistical Analysis, Time-Series Analysis, A/B Testing
Model Evaluation	MAE, MSE, RMSE, R^2, Precision, Recall, F1, Cross-Validation, Silhouette Score, Elbow Method
LLMs & NLP	Groq API, LLaMA 3, Prompt Engineering, Semantic Parsing, Unstructured Data Pipelines, Multi-Agent Orchestration
Deployment	FastAPI, Streamlit, Vercel, Uvicorn, Joblib, SQLAlchemy
Visualization	Matplotlib, Seaborn, Plotly, Streamlit
Tools	Git, GitHub, Jupyter Notebook, VS Code, Google Sheets

Education

SJCET Palai — *B.Tech in AI and Data Science* *Sept 2023 – Present*
Kottayam, Kerala, India

- CGPA: 7.31 / 10
- Relevant Coursework: Artificial Intelligence, Artificial Neural Networks, Machine Learning, Deep Learning, Big Data Analytics, Data Science, Database Management Systems, Data Structures & Algorithms, Probability & Statistics

Work Experience

British Airways — *Data Science Virtual Experience* *February 2026*
Remote (Forage)

- Surfaced actionable customer satisfaction drivers by scraping and performing EDA on 1,000+ unstructured review records, identifying sentiment patterns across key service dimensions
- Boosted booking behavior prediction by engineering domain-specific features and training a Scikit-learn classification model; documented findings and delivered results in a structured executive-facing presentation

Commonwealth Bank — *Introduction to Data Science Virtual Experience* *February 2025*
Remote (Forage)

- Reduced downstream data errors by designing and implementing end-to-end Python and SQL pipelines for ingestion, cleaning, and transformation of large-scale structured financial datasets
- Ensured data governance compliance by applying anonymization techniques to sensitive customer records across multi-table relational datasets
- Accelerated reporting turnaround by designing relational schemas and SQL-based analytical workflows optimized for large-scale querying and dashboard-ready output

Projects

LeafNetwork — Multi-Agent AI Platform | github.com/shyamhari1074/The_Leaf_Network
Tech: React 19 (Vite), FastAPI, PyTorch, HuggingFace (MobileNet V2), Groq API (LLaMA 3), SQLAlchemy, SQLite, Vercel

- Designed and deployed a full-stack multi-agent AI platform to Vercel, integrating a finetuned MobileNet V2 Vision Agent for real-time crop disease detection, served via FastAPI and consumed by a React 19 frontend — end-to-end ML system ownership from model training to production deployment
- Engineered a multi-agent orchestration layer (Vision, Climate, Satellite agents) synthesizing outputs from 3 specialized ML models and external APIs into unified crop recommendations, demonstrating real-world ML system architecture at scale
- Integrated Groq API (LLaMA 3) with prompt engineering workflows for a context-aware AI assistant, and a Data.gov pricing API with trend analysis algorithms delivering real-time Buy/Sell/Hold market recommendations

Reinforcement Learning for Time-Series Financial Optimization
Tech: Python, Stable-Baselines3 (PPO), PyTorch, Gymnasium

- Outperformed baseline statistical hedging methods by ∼80% by designing and training a PPO reinforcement learning agent on sequential financial time-series data with custom reward functions and domain-specific state feature engineering
- Benchmarked RL policy performance against traditional statistical approaches, documenting experimental methodology and results for reproducibility

Customer Segmentation & Churn Prediction Platform
Tech: Python, Scikit-learn (K-Means, Logistic Regression), Pandas, NumPy, Streamlit, Joblib

- Deployed an end-to-end ML pipeline combining K-Means clustering (5 behavioral segments, validated via Elbow Method and Silhouette Scoring) and Logistic Regression churn classification, with Joblib model serialization for production inference
- Improved classifier reliability by calibrating evaluation metrics (Precision, Recall, F1) for class imbalance; shipped a real-time Streamlit inference dashboard enabling live predictions from user-input data

Padik AI — Automated LLM Data Pipeline
Tech: Python, LLaMA 3.1 (Groq API), YouTubeTranscriptApi, Prompt Engineering

- Eliminated manual data processing by building an agentic NLP pipeline that transforms raw unstructured YouTube transcript data into structured outputs (study notes, Q&A pairs) using LLaMA 3.1 via Groq API — zero manual processing post-deployment
- Achieved deterministic structured outputs by engineering and iterating on prompt workflows for semantic parsing of noisy, incomplete real-world text at scale

E-commerce Customer Behavior Analysis & Predictive Modeling
Tech: Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

- Identified primary revenue drivers by conducting EDA and feature analysis on 2,000+ customer records; built and validated a Linear Regression model (MAE, MSE, RMSE, R^2, residual normality checks) with a 70/30 train-test split
- Quantified coefficient-level business insight: membership duration, app time, and session length drove $61.27, $38.59, and $25.98 annual spend increases per unit respectively, directly informing product investment

priorities

Autonomous Game Agent via Neuroevolution
Tech: Python, PyTorch, Genetic Algorithms

- Designed and trained a lightweight PyTorch neural network agent using genetic algorithms as a gradient-free optimization alternative to backpropagation, implementing iterative fitness-based selection across generations for real-time decision-making without supervised labels

Certifications & Virtual Experience

- British Airways Data Science — Forage Virtual Experience Program (February 2026)
- Commonwealth Bank Introduction to Data Science — Forage Virtual Experience Program (February 2025)

Achievements

- Deployed LeafNetwork — a production multi-agent ML platform — to Vercel, integrating 3 AI agents, 3 external APIs, a PyTorch vision model, FastAPI backend, and React 19 frontend, built and shipped independently
- Outperformed baseline statistical financial methods by ∼80% via a custom PPO reinforcement learning agent with engineered reward functions and state feature design
- Built and deployed two end-to-end ML systems to production (LeafNetwork on Vercel, Segmentation platform on Streamlit) demonstrating full ML lifecycle ownership from data to deployment
- Completed two industry virtual experience programs (British Airways, Commonwealth Bank) applying ML and data science to real-world business problems

Languages

English (Fluent) | Hindi | Malayalam