

Swayam Shah

+91-7434946603 | swayamshah0406@gmail.com | LinkedIn | GitHub | Kaggle

Final-year ECE undergraduate specializing in LLM applications, agentic AI pipelines, and data engineering translating unstructured data into business decisions through scalable Python backends and NLP systems.

TECHNICAL SKILLS

Languages & Frameworks: Python (OOP, Asyncio), C++, SQL (PostgreSQL), FastAPI, Node.js, Express.js, REST APIs

AI / ML & LLM Engineering: LLMs, RAG, Agentic AI, LangChain (LCEL), LangGraph, OpenAI API, Gemini API, FAISS, ChromaDB, Vector Embeddings, NLP, Prompt Engineering

Data Engineering: ETL Pipelines, Data Wrangling, Pandas, Apache Spark, Apache Airflow, Sentiment Analysis, Structured & Unstructured Data, Business Intelligence

Tools & Platforms: DataBricks, React.js, MongoDB, Socket.io, Docker, PostgreSQL, SQLAlchemy, Git/GitHub, Railway/Vercel (Cloud Deployment)

EDUCATION

Indian Institute of Information Technology (IIIT), Una
B.Tech in Electronics and Communication Engineering (ECE)

2023 – 2027
CGPA: 7.69

Prakash Higher Secondary School, Ahmedabad
Class XII, CBSE

2023
89%

PROFESSIONAL EXPERIENCE

Data Engineer intern

Hem Infotech

May 2025 – Present

Ahmedabad, Gujarat

- **LLM-Powered ETL Pipeline:** Built an **OpenAI API**-driven pipeline to automate lead classification and scoring across 1,300+ client records — replacing a fully manual process with a scalable, production-ready workflow.
- **Data Engineering:** Designed end-to-end **ETL workflows** using **Pandas** and **SQL** to merge, clean, and transform disparate client datasets into a reliable single source of truth for downstream business analysis.
- **AI-Augmented Decision Support:** Developed an automated reporting pipeline where **LLM-generated summaries** translated structured data into actionable insights, enabling faster, evidence-based decisions for sales leadership; maintained the codebase collaboratively via **Git/GitHub**.

TECHNICAL PROJECTS

PriceIQ — Autonomous Pricing Intelligence Engine

Python, LangGraph, LangChain, FastAPI, PostgreSQL, Groq (Llama 3.3 70B), Firecrawl, Next.js, Docker

- **Agentic AI Pipeline:** Built a **LangGraph** 4-node multi-agent system that scrapes competitor prices via **Firecrawl**, computes margin analytics via **Pandas**, and generates **LLM-reasoned** pricing recommendations — replacing manual competitor tracking entirely.
- **NLP Market Signals:** Integrated **VADER sentiment analysis** on competitor reviews as a real-time pricing signal, combining NLP-derived intelligence with structured sales data to produce grounded, business-justified recommendations.
- **Production Deployment:** Implemented a **Groq → Gemini → rule-based** fallback chain with **APScheduler** automation; containerized via **Docker** and deployed on **Railway + Vercel** cloud platforms with zero-downtime graceful degradation.

RetailLake — Open-Source Retail Data Lakehouse Platform

Apache Spark, Delta Lake, dbt, DuckDB, Redpanda, MinIO, Great Expectations, Airflow, Streamlit, Terraform

- **Lakehouse Architecture:** Engineered an end-to-end lakehouse ingesting **50,000+ retail transactions** via batch plus live **click-stream events** via streaming, using **MinIO** (S3-compatible) and **Redpanda** (Kafka-compatible) as fully open-source, single-command-launchable infrastructure.
- **Distributed Processing & Warehouse Modeling:** Processed data with **Apache Spark + Delta Lake** for ACID transactions and time travel, then built a **dbt + DuckDB** star schema warehouse (fact_sales, dim_product, dim_date) directly from Delta Parquet files.
- **Workflow Orchestration & Data Quality:** Scheduled and monitored automated data workflows using a production-style **Apache Airflow** DAG with retries and `max_active_runs=1`, ensuring timely and accurate data processing; enforced **22 automated data quality checks** via **Great Expectations** that halt the pipeline on failure, surfaced insights via a **Streamlit** dashboard, with **Terraform** provisioning all infrastructure declaratively.

ACHIEVEMENTS & EXTRACURRICULARS

Competitive Programming: Highest LeetCode contest rating of **1599**, reflecting strong DSA and problem-solving ability.

Academic Excellence: Secured a **96.86 percentile** in JEE Mains and qualified for JEE Advanced, placing me among the top engineering entrance candidates nationwide.

Leadership: Core Member of Aavesh Electronics Club at IIIT Una; coordinated technical events for the annual college fest, Meraki.