TOHID SHAIKH

+91-9510944574 | shaikhtohid921@gmail.com | linkedin.com/in/shaikh-tohid | github.com/T786-eng | Dahod, Gujarat, India

PROFESSIONAL SUMMARY

Backend and ML engineer with hands-on experience building distributed systems, LLM-integrated pipelines, and data-intensive applications. Developed 6 end-to-end projects spanning encrypted payment networks, custom vector databases, and network inspection engines — with measurable outcomes across datasets exceeding 5M records.

TECHNICAL SKILLS

Languages: Python, Java, SQL, Bash

Core CS: Data Structures & Algorithms, System Design, Scalable Architecture

Backend & Systems: REST APIs, Flask, SQLAlchemy, Multithreading, Distributed Systems, TCP/IP

AI / LLM: LLM Integration, RAG Pipelines, Vector Databases, Anthropic Claude API, Prompt Engineering

ML & Data: Scikit-learn, Pandas, NumPy, TF-IDF, SMOTE, Random Forest, NLP

Tools: Git, Streamlit, Jupyter, Plotly, Matplotlib

EXPERIENCE

ML Engineer Intern Jan 2026 – Feb 2026
Edunet Foundation (IBM SkillsBuild) *Gujarat, India*

- Built a Random Forest classification pipeline on 12K records, achieving **91.2% accuracy** and **0.87 F1-score** through structured feature engineering and cross-validated model selection.
- Designed modular preprocessing workflows using Pandas and Scikit-learn, eliminating data leakage and ensuring reproducibility across training and evaluation environments.

ML Engineer Intern Dec 2025 – Jan 2026
3Skill *Gujarat, India*

- Developed a fraud detection model on **20K+ imbalanced financial transactions**; applied SMOTE oversampling to improve recall from **61% to 84%**, reaching **92% accuracy** and **0.89 ROC-AUC**.
- Reduced false-negative rate by **38%** through targeted feature selection and threshold tuning on skewed real-world transaction data.

PROJECTS

UPI Offline Mesh | *Python, Flask, SQLAlchemy, RSA-OAEP, AES-GCM, Multithreading* GitHub

- Built an offline UPI payment system using a gossip-based mesh network; implemented **hybrid RSA-OAEP + AES-256-GCM** encryption to secure payment packets relayed across virtual bridge nodes.
- Designed a thread-safe idempotency cache with **optimistic locking** (SQLAlchemy versioning) to prevent duplicate settlements under concurrent multi-bridge packet delivery.

Vector Database with RAG Pipeline | *Python, Flask, HNSW, KD-Tree, Ollama, REST API* GitHub

- Implemented a vector database from scratch with **HNSW**, KD-Tree, and Brute-Force search; achieved **sub-200ms** approximate nearest-neighbor query latency across high-dimensional embeddings.
- Built a RAG pipeline using Ollama (nomic-embed-text + llama3.2) for document-grounded question answering with semantic chunking and real-time concurrent retrieval.

UIDAI Aadhaar Hackathon — Project DRAM | *Python, Pandas, SciPy, Plotly, Streamlit* GitHub | Live Demo

- Processed **5M+ Aadhaar records** across 12 data sources; classified **807 districts** using Z-score anomaly detection and a custom Updates-to-Enrolment Ratio (UER) metric into 3 infrastructure priority tiers.
- Improved regional query throughput by **10x** using vectorized boolean masking; presented findings via live Streamlit dashboard at UIDAI National Innovation Challenge 2026.

Deep Packet Inspection Engine | *Python, Multithreading, TCP/IP, TLS, PCAP* GitHub

- Developed a 10-module DPI engine that parses raw PCAP files across Ethernet/IP/TCP/UDP layers; identified **20+ applications** via TLS SNI fingerprint matching with stateful 5-tuple flow tracking.
- Implemented a multithreaded Reader $\rightarrow$ Load Balancer $\rightarrow$ Fast Path pipeline with consistent hashing to maintain per-connection affinity across parallel processing threads.

EDUCATION

Government Engineering College, Dahod Jun 2026
Bachelor of Technology, Electrical Engineering *Dahod, Gujarat*

CERTIFICATIONS

Google – AI & ML on Google Cloud | Google – Vertex AI Prompt Design | IBM – AI Fundamentals | Cisco – Python Essentials 1 & 2 | McKinsey.org – Forward Program