Ayush Patel *Machine Learning Engineer*

work.ayushpatel18@gmail.com +91 8200880686 LinkedIn GitHub

SUMMARY

Machine Learning Engineer with experience leading AI projects from design to production, specializing in RAG, LLM fine-tuning (PEFT/LoRA), and vector database integration (Weaviate, Pinecone). Skilled in building and deploying scalable backend systems on cloud platforms. Strong background in NLP, deep learning, and model optimization, with a focus on efficiency, accuracy, and delivering production-ready AI solutions.

EDUCATION

Master's in Data Science, Jul 2023 – May 2025 | Gandhinagar, Gujarat
Dhirubhai Ambani Institute of Information and Communication Technology
(8.54 CGPA)

Bachelor's in Statistics, *St. Xavier's College* Jul 2020 – Apr 2023 | Ahmedabad, Gujarat
(7.74 CGPA)

WORK EXPERIENCE

F(x) Data Labs, *Associate Machine Learning Engineer* Dec 2024 – Present | Ahmedabad, Gujarat

Automated Form Generation (Project Lead) | *OpenAI, LangChain, Weaviate, Docker*
- Engineered a production-grade AI backend using Python, FastAPI, Weaviate, and OpenAI GPT (RAG) to deliver sub-5-second, context-aware semantic search and automated form generation across multi-format documents.
- Boosted operational efficiency by 80% through hybrid semantic search, vector embeddings, and automated validation—achieving 95% accuracy and reducing manual form preparation time from hours to minutes.
- Architected and deployed a containerized, cloud-ready solution with Docker and Docker Compose, integrating document parsing, robust logging, and error-handling for enterprise-grade reliability.

Numerology Chatbot | *Pinecone, MySQL, AWS*
- Built an AI-driven numerology chatbot using OpenAI and LangChain with prompt-engineered query classification, semantic retrieval via Pinecone, and tone-adaptive responses tailored to user profiles.
- Designed numerology-based embedding logic and feedback refinement, improving NLP coherence, personalization, and scalable insight generation across diverse range of user intents.
- Deployed the system on AWS in collaboration with the FullStack team for API integration, reducing latency from 15–20s to 5–7s and improving response quality by 35% through optimized vector retrieval and prompt workflows.

Customer Lifetime Value (CLV) Prediction | *Python, SQL, Power BI, Google Gemini LLM*
- Developed a predictive model using Random Forest algorithm to identify high-value customers from 9,000+ insurance records, enabling targeted retention strategies and data-driven marketing decisions.
- Created interactive Power BI dashboards and comprehensive SQL analysis with 200+ queries covering customer segmentation, retention patterns, and sales channel performance for executive reporting.
- Integrated Google Gemini LLM with MySQL database to build natural language Q&A system, empowering non-technical users to generate complex SQL queries and extract business insights through conversational interface.

ACADEMIC PROJECTS

AI Tutor, | *RAG, PEFT, LoRA, LLaMA 3*
- Built an AI Tutor pipeline simulating expert-led teaching by extracting YouTube transcripts, generating context-rich Q&A pairs using RAG, and fine-tuning LLaMA-3 via PEFT with LoRA.
- Delivered structured, high-quality responses with 1.05 perplexity, 0.55 BLEU, 0.74 ROUGE-L, and <4.3s average latency.

End-to-End Fashion Item Classification using CNN, | *DL, TensorFlow, FastAPI, Docker*
- Developed a FastAPI-based REST API serving a Keras CNN model trained on the Fashion MNIST dataset containing 70K images & 10 classes, achieving 88% test accuracy.
- Trained the model and deployed near real-time inference with an average inference latency of <4-5s per image.
- Containerized the application with Docker and deployed on AWS EC2, reducing deployment time by 60% and ensuring scalable, production-grade performance.

SKILLS & TOOLS

Technical Skills: ML, DL, Data Analysis, NLP, Statistical Analysis, Information Retrieval, Fine-tuning LLMs, RAG

Programming Languages: Python, SQL

Tools: VS Code, PostgreSQL, Git, GitHub, Google Colab, Tableau, Excel, Cursor, Windsurf

Technologies/Frameworks: Scikit-learn, TensorFlow, PyTorch, Keras, LangChain, FastAPI, Streamlit, VectorDB (FAISS, Pinecone, Weaviate), AWS

Soft Skills: Cross-Team Collaboration, Effective Communication, Fast Learner, Continuous Upskilling