

Farhan Shaikh

+91-7096362686 | farhanshaikh11092002@gmail.com | LinkedIn | GitHub | farhan11.xyz

EDUCATION

NMIMS University

Bachelor of Technology in Information Technology; CGPA: 3.42/4.00

Mumbai, India

Aug 2024

PROFESSIONAL EXPERIENCE

NN & Sons

AI/ML Engineer

Silvassa, Dadra and Nagar Haveli

Oct 2025 – Ongoing

- Engineered end-to-end multilingual document translation system using llama-4-maverick-17b-instruct via Groq API and Streamlit, enabling bidirectional translation between Indian regional languages for PDF and DOCX files, saving 500+ hours of manual work
- Designed modular NLP pipeline with document parsing, token-safe chunking, LLM-based translation, and structure-preserving reconstruction to maintain formatting integrity across 8+ regional languages
- Built production RAG chatbot using LangChain and ChromaDB to answer employee queries on printing machine troubleshooting, reducing support ticket resolution time by implementing vector similarity search and context-aware response generation

Ernst & Young

Senior Analyst

Mumbai, India

July 2024 – Sept 2025

- Conducted independent data collection, validation and integrity analysis for Fortune 500 clients, ensuring risk readiness and identifying control gaps aligned with regulatory compliance frameworks
- Executed technology risk and control testing across global telecommunications and financial clients, specializing in ITGC, IT application controls, and IPE assessment within SAP GRC and SAP S/4HANA environments, AI applications and custom-built financial systems
- Delivered actionable technical insights via dashboards using Power BI & Python libraries, translating complex data into business-focused recommendations for executive leadership

TECHNICAL PROJECTS

Parameter-Efficient LLM Fine-Tuning | PyTorch, Hugging Face, LoRA, Unsloth

- Implemented PEFT pipeline for LLaMA-3.2-1B-Instruct using LoRA adapters on attention layers, reducing trainable parameters by 99% while maintaining model performance through quantization-aware training
- Leveraged Unsloth framework for optimized kernel execution, achieving 40% faster training and 60% reduced VRAM consumption on resource-constrained GPUs

Kidney Tumor Classification System | TensorFlow, MLOps, Docker, AWS

- Developed end-to-end CNN-based medical imaging pipeline achieving 88.26% validation accuracy under hardware constraints, implementing modular MLOps architecture with DVC for versioning and MLflow for experiment tracking
- Deployed production Flask API containerized with Docker and automated CI/CD via GitHub Actions, hosted on AWS EC2 with ECR for scalable real-time inference

Credit Default Payment Classifier | Scikit-Learn, XGBoost, Random Forest

- Built ML-based credit risk assessment system for financial institutions, evaluating multiple algorithms including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting
- Achieved 93.84% accuracy and 0.86 AUC score using hyperparameter-tuned XGBoost classifier through comprehensive EDA, feature engineering, and model optimization techniques

TECHNICAL SKILLS

Languages: Python, SQL, HTML, CSS, JavaScript

ML/AI Frameworks: Pandas, NumPy, Matplotlib, Plotly, PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, LangChain, Hugging Face Transformers

Tools & Technologies: Docker, AWS (EC2, ECR, IAM), Git, MLflow, DVC, FastAPI, Tableau, Excel, MySQL, Linux
Specializations: Statistical Analysis, Feature Engineering, EDA, A/B Testing, Large Language Models, Deep Learning, Computer Vision, NLP, MLOps, RAG Systems

Certifications: Google Advanced Data Analytics, HackerRank 5 star SQL, IBM Data Science, McKinsey Forward