

FARHAN SHAIKH

Website: farhan11.xyz | Ph no: +91-7096362686 | farhanshaikh11092002@gmail.com | [LinkedIn](#) | [Github](#)

ACADEMICS

NMIMS University, Mumbai	August 2024
BTECH, Information Technology (CGPA: 3.42 / 4.00)	
Lions English School, Dadra & Nagar Haveli	June 2020
PCM Higher Secondary- CBSE (85.6%)	
Lions English School, Dadra & Nagar Haveli	June 2018
Secondary Education- CBSE (92.6%)	

TECHNICAL SKILLS

Languages: Python, SQL
Skills: Data analytics, Machine Learning, Deep Learning, Computer vision, SQL, Object Oriented Programming, Git, NLP
Libraries/Frameworks: Numpy, Pandas, Matplotlib, Plotly, Sci-kit learn, PyTorch, Torchvision, Tensorflow, keras, mlflow, Transformers, LLM, Langchain
Tools: MySQL, Windows, Linux, MS excel, Google-Collab, Jupyter-Notebook, VS Code, Cursor, Tableau, Github, MS Office, Docker, AWS (IAM, ECR, EC2)
Certifications/Badges: Google Advanced Data Analytics, HackerRank 5 star SQL, IBM Data Science 101, McKinsey Forward Program

PROJECTS

Parameter-Efficient Instruction Tuning of LLM (LLM, Hugging Face, Unslot, [Notebook Link](#))

- Designed and implemented a parameter-efficient fine-tuning (**PEFT**) pipeline for the open-source **LLaMA-3.2-1B-Instruct** model using an Alpaca-format instruction dataset, enabling effective instruction-following adaptation while minimizing GPU memory usage and training cost.
- Applied **LoRA**-based fine-tuning by freezing base model weights and injecting low-rank adapters into attention and projection layers, allowing task-specific learning with only a small subset of trainable parameters compared to full fine-tuning.
- Integrated quantization-aware model loading and leveraged the **Unslot** framework to utilize optimized kernels and memory-efficient training routines, achieving faster training and reduced VRAM consumption while maintaining model performance.

New York City Taxi Fare Prediction (Scikit-Learn, XGboost Regressor, [Notebook Link](#))

- Developed as part of a Google playground competition hosted by **Google Cloud** on Kaggle, focuses on predicting New York City taxi fares based on pickup and drop-off locations.
- Performed **extensive data cleaning, preprocessing** including handling missing values, applying **Min-Max** scaling, and using **One-Hot Encoding** for categorical variables and **feature engineering**, including **time-based** and **geospatial features** (using the Haversine formula for distance) multiple machine learning models like tree-based models, and ensemble models like Random Forest Regressor, Gradient boosting were evaluated.
- After hyperparameter tuning the **XGBoost** Regressor outperformed all achieving an **RMSE of 3.25** on Kaggle, placing the model in the **top 30%** of the Kaggle leaderboard.

Kidney Tumor Classification System (Deep Learning, Computer vision, MLOps, [Github Link](#))

- Addressed the challenge of automated **kidney tumour** detection from medical images by building an end-to-end **CNN**-based classification system to assist accurate and **timely diagnosis**, achieving **88.26%** validation accuracy under **hardware constraints**.
- Implemented a modular **MLOps** pipeline using **TensorFlow** and **Keras** with image preprocessing, regularization, and structured stages for data ingestion, validation, training, evaluation, and prediction, ensuring reproducibility via DVC and MLflow (DAGsHub).
- Deployed a production-ready Flask application containerized with Docker, automated through GitHub Actions CI/CD, and hosted on AWS (ECR, EC2) to enable scalable, real-time inference.

Credit Default Payment Classifier (Scikit-Learn, Random Forest Classifier, XGboost Classifier, [Notebook Link](#))

- This project develops a machine learning based **credit default** prediction system to help **financial institutions** assess **borrower risk**.
- Multiple models were evaluated, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. After data cleaning, preprocessing, Exploratory Data analysis, and Hyperparameter tuning, the **XGBoost** classifier achieved the best results, reaching **93.84%** accuracy and an **AUC score of 0.86**.

PROFESSIONAL EXPERIENCE

NN & Sons - Applied AI/ML (intern) (Sept 2025- Jan 2026)

- Developed an end-to-end multilingual document translation system using the **llama-4-maverick-17b-instruct model** via the Groq API and **Streamlit** UI, enabling bidirectional translation between Indian regional languages for both PDF and DOCX files
 - Implemented a modular **NLP** pipeline that extracts text from documents, segments content into safe token-sized chunks, translates each segment using an **LLM**, and reconstructs the translated output while preserving the original document structure, saving over 500 hours of manual work.
- Built a **RAG chatbot** to answer employee queries related to printing machine errors, troubleshooting steps, and solutions, using product manuals and technical documentation as the knowledge source. used **LangChain** and **ChromaDB** that retrieves relevant context from a **vector database** and generates accurate responses using a large language model. Implemented document loading, **text chunking, embedding creation**, similarity search, and prompt-based response generation to improve factual accuracy and reduce hallucinations.

Ernst & Young - Senior Analyst (full-time) (July 2024- Sept 2025)

- Executed Technology Risk and control testing engagements across global clients, focusing on ITGC, IT application controls, and IPE assessment, including work for one of the world's largest telecommunications organizations.
- Performed independent **analysis** and **validation** of **client data**, ensuring data integrity and audit readiness prior to control testing and reporting.
- Conducted in-depth control reviews within **SAP GRC**, **SAP S/4HANA**, and custom-built **financial systems**, identifying control gaps and recommending remediation aligned with regulatory and compliance frameworks.
- Delivered clear, actionable insights through client presentations and formal documentation, translating technical findings for both technical and business stakeholders.

Ernst & Young - Technology Consulting (Intern) (Jan 2024- July 2024)

- Gained hands-on exposure to **Global risk methodologies, risk frameworks, and enterprise systems**, building a strong foundation in technology risk and compliance.
- Supported ongoing **Technology Risk assessments** by assisting engagement teams in control walkthroughs and documentation for telecom and financial services clients.
- Contributed to **ITGC, ITAC, and IPE evaluations** by gathering evidence, preparing documentations, and validating control design and implementation.