Raheel Abdul Rehman

+91 8147126568 , raheelrehman312001@gmail.com , www.linkedin.com/in/raheelrehman/ , https://raheel-rehman.vercel.app/ , https://huggingface.co/Raheel31 , https://github.com/Raheel31

Summary

Machine Learning & Data Science Engineer skilled in building end-to-end ML projects through hands-on work in model development, data pipelines, and production-ready Python systems. ML expertise strengthened through applied projects and professional coursework. Industry background at Goldman Sachs as a Data and Strategic Solutions Analyst, building automated ETL (Extract, Transform, Load) systems, analytics dashboards, and data products used across global business units.

Skills

- **Programming:** Python, SQL, R, C
- **Machine Learning:** PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, LightGBM, HuggingFace Transformers, LangChain, LangGraph, Optuna, FAISS, ChromaDB,OpenCV, Retrieval-Augmented Generation (RAG)
- **Data Analysis:** Pandas, NumPy, Matplotlib, Seaborn
- **Data Engineering:** PySpark, MongoDB, BigQuery
- **Cloud & DevOps:** AWS, Docker, Git, GitHub, GitLab, Pytest, Pylint, CI/CD
- **Frameworks & Tools:** FastAPI, Flask, Streamlit, Gradio, Requests, BeautifulSoup, Postman, Prettify
- **BI Tools:** Tableau, Power BI, Excel, Alteryx, IBM Watson Studio, Aqua Data Studio

Experience

Goldman Sachs - Data Analytics and Strategic Solutions **Jan 2023 - Jul 2025**

- Built a data-quality monitoring dashboard using Python, SQL, Alteryx, and Tableau, improving data-integrity detection to 99% accuracy, enabling proactive regulatory-risk mitigation across global teams.
- Developed a real-time exposure-tracking system using SQL and Tableau, enhancing APAC risk visibility for $2B+ collateral positions and supporting compliance workflows.
- Automated ETL workflows using Python and Alteryx, reducing manual processing and improving pipeline performance by 90%, enabling faster analytics delivery across EMEA/APAC.
- Designed interactive revenue and user-analytics dashboards in Tableau and SQL, enabling real-time monitoring across 2M+ records and improving operational reporting efficiency.

Shinkan Pvt. Ltd. - Data Scientist Intern **Jan 2022 – Jul 2022**

- Developed a supervised audio-classification model using Scikit-learn and Keras, achieving 98% accuracy across 1,500+ audio samples, improving automated speech recognition workflows.
- Processed raw audio data using NumPy and Pandas, cleaning and structuring features to boost downstream ML model performance and reliability.
- Visualized key acoustic features using Matplotlib and Seaborn, identifying influential variables and improving model interpretability.

Education

Manipal Institute of Technology

Bachelor of Technology in Mechatronics, Minor Specialization Data Science

Certifications

- IBM Data Science Professional Certificate
- IBM Deep Learning with PyTorch, Keras and TensorFlow Professional Certificate
- IBM Generative AI Engineering Professional Certificate

Projects

GenAI Career Assistant Chatbot

- Built an AI job-fit analysis system using Gradio, FAISS, and Hugging Face API to parse resumes, retrieve embeddings, and generate structured LLM insights.
- Implemented a lightweight RAG pipeline using MiniLM + FAISS to produce job-fit scores, strengths, missing skills, and actionable recommendations

Synthetic Stock Data Generator

- Built a hybrid Autoencoder– Generative Adversarial Network (GAN) generative model to produce realistic synthetic OHLCV stock-market sequences, with Gradio dashboards for real-vs-synthetic comparison and interactive visualization.
- Demonstrated utility via a downstream classifier where synthetic augmentation improved validation accuracy by 6% and reduced loss by 12%.

Guitar Recommendation System

- Engineered a full-stack ML pipeline using Principal Component Analysis (PCA), MiniBatch K-Means, and cosine similarity to recommend skill-appropriate songs based on chord difficulty, tempo, and style.
- Deployed a FastAPI + Docker backend with a custom frontend, delivering top-5 song recommendations and difficulty-balanced practice exercises via efficient vector search.