

FAHEEM KHAN

Aligarh, Uttar Pradesh | +91-8279575494 | faheemthakur23@gmail.com | Portfolio | LinkedIn | GitHub

SUMMARY

Data-driven professional specializing in Generative AI, with 1.5+ years of internship experience. Proven ability to build end-to-end AI solutions, from creating scalable ETL pipelines on GCP to finetuning and deploying LLMs. Skilled in Python, SQL, and frameworks like LangChain and Hugging Face to process both structured and unstructured data, enhance model performance, and automate data-driven workflows.

TECHNICAL SKILLS

- Programming:** Python (Pandas, NumPy), Git, GitHub, GitHub Actions
- Data Engineering:** SQL (MySQL, PostgreSQL), ETL, Data Validation, BigQuery, Mage AI
- Visualization:** Looker Studio, Matplotlib, Seaborn
- Cloud & Tools:** GCP, Docker, Technical Documentation
- Gen AI:** LLMs, Rag, Vector databases, Langchain

EXPERIENCE

- Machine Learning Intern** | Asan Innovators, Freelance Jan 2025 – Present
- Developed automated data pipelines using Python for extracting, transforming, and loading datasets from multiple sources.
  - Deployed data services via Flask APIs, enabling real-time data access for downstream analytics.
  - Containerized data applications using Docker to streamline deployment and ensure consistent environments.
  - Maintained clear technical documentation for all ETL processes, reducing onboarding time and enhancing cross-team collaboration.
- Data Engineer Intern** | Everlytics Data Science Pvt Ltd Mar 2024 – Sep 2024
- Designed and built ETL workflows in Python and SQL to process structured (CSV) and semi-structured (JSON) data.
  - Performed extensive data validation and cleaning, boosting model training efficiency and reliability.
  - Optimized SQL queries for large datasets, improving query performance by 30%.
  - Collaborated with analysts to understand reporting requirements and build reusable data transformation scripts.
- Data Science Intern** | Postulate Info Tech Pvt Ltd Jan 2023 – Apr 2023
- Conducted feature engineering using Pandas to improve the accuracy of business metric predictions by 20%.
  - Wrote and optimized SQL queries to retrieve, join, and aggregate data from relational databases.
  - Developed and documented 20+ data reports and notebooks to support stakeholder insights and decision-making.
  - Assisted in exploratory data analysis to uncover patterns and trends for predictive modeling tasks.

EDUCATION

Maulana Azad National Urdu University	<b>Master of Computer Applications</b>   CGPA: 8.57	Oct 2022 – May 2024
Dr. B R Ambedkar University	<b>Bachelor of Computer Applications</b>   CGPA: 7.0	Aug 2018 – Oct 2021

PROJECTS

- AI-Powered Medical Q&A System** | Python, LangChain, Gemini, Pinecone, Docker, GCP July 2025
- Engineered a RAG system processing a 600+ page medical text, creating over 5,800 vector embeddings to ensure fact-based, low-latency responses.
  - Implemented a history-aware retrieval chain to maintain conversational context, improving follow-up question accuracy and user experience.
  - Automated the end-to-end deployment pipeline to Google Cloud using GitHub Actions, reducing manual deployment time by 100% and ensuring consistent builds.
  - Live Demo
- Hotel Reservation Cancellation Prediction** | Python, Scikit-learn, Flask, Docker January 2025
- Built supervised learning model achieving 88% accuracy using scikit-learn on 10,000+ records dataset.
  - Implemented comprehensive data preprocessing and feature engineering pipeline to improve model performance.
  - Deployed ML model using Flask API with Docker containerization and CI/CD pipeline integration.
  - Live Demo

CERTIFICATIONS

- MLOps on GCP (CI/CD, Kubernetes, Jenkins), 2025
- Data Science Certificate, Pro5 (Top 5%), 2024