Khuzaima Hassan

AI/ML Engineer | Data Science Undergraduate | Python Developer

Karachi, Pakistan | khuzaimahassan52@gmail.com | +92 321 3937589 | linkedin.com/in/khuzaima-hassan-2004mkh | github.com/KhuzaimaHassan

PROFESSIONAL SUMMARY

Data Science undergraduate at NED University of Engineering & Technology with hands-on experience in machine learning and Generative AI, gained through a Data Science Internship at 10Pearls. Proficient in Python, SQL, FastAPI, and Streamlit, with practical experience integrating LLM APIs (Gemini, Groq) for real-time data Q&A, automated reporting, and interactive dashboards. Strong foundation in classification, regression, feature engineering, model evaluation, and end-to-end ML pipeline design, with a track record of turning raw data into production-style applications.

TECHNICAL SKILLS

Programming: Python, SQL

Machine Learning: Classification, Regression, Feature Engineering, Data Preprocessing, Model Evaluation, Cross-Validation, Model Monitoring, Pipeline Design

Generative AI & LLMs: Gemini API, Groq API, LLM Integration, Prompt Engineering, Retrieval-Augmented Q&A, GenAI App Development

Frameworks & Tools: FastAPI, Streamlit, Git, GitHub, Jupyter Notebook, Hugging Face

Databases & Visualization: PostgreSQL, MySQL, Power BI, Pandas, NumPy, Matplotlib, Seaborn

PROFESSIONAL EXPERIENCE

10Pearls — Data Science Intern *Dec 2025 – Feb 2026*

- Built an end-to-end ML pipeline for an AQI prediction system, covering data collection, cleaning, preprocessing, and feature engineering using Python, Pandas, and scikit-learn.
- Trained and evaluated regression and classification models to forecast AQI levels, applying cross-validation to improve generalization.
- Developed and deployed an interactive dashboard to visualize AQI predictions and environmental insights, supporting data-driven monitoring decisions.

Japanese Language & IT Internship (Plus W × OEC) — Internee *Apr 2025 – Sep 2025*

- Worked on database-related tasks to support backend development, including adding unique constraints to database tables to ensure data integrity and prevent duplicate records.
- Wrote and executed database migration scripts to safely update existing tables and maintain data consistency across environments.

PROJECTS

Risk Scoring Platform — End-to-End ML System

GitHub: github.com/KhuzaimaHassan/risk-scoring-platform

- Built an end-to-end fraud risk scoring platform using Python, PostgreSQL, and FastAPI, covering data ingestion, feature engineering, and model training.
- Implemented model versioning to manage multiple model iterations, enabling reproducible experimentation and safe rollback.
- Designed a model monitoring layer with statistical drift detection to track prediction distributions and flag performance degradation.

Insight Agent — LLM-Powered Data Analysis App

GitHub: github.com/KhuzaimaHassan/insight-agent

- Integrated Google's Gemini API to build an LLM-powered application that answers natural language queries over user-uploaded datasets in real time.
- Applied prompt engineering to ground LLM responses in uploaded data, enabling accurate, context-aware question answering.
- Built a Streamlit dashboard with automated dataset summaries, visualizations, and an interactive Q&A interface for non-technical users.

AI Academic Mentor — GenAI Hackathon Project

GitHub: github.com/KhuzaimaHassan/ai-academic-mentor

- Built a GenAI tool using the Groq API to generate personalized, performance-based academic reports for students.
- Designed prompt templates that convert raw student performance data into clear, actionable feedback and recommendations.
- Developed a Streamlit interface to display student data alongside auto-generated mentor reports from a Data Science Hackathon.

Real-Time Fraud Detection System — Machine Learning

GitHub: github.com/KhuzaimaHassan/real-time-fraud-detection

- Addressed severe class imbalance in fraud data using SMOTE (Synthetic Minority Over-sampling Technique) for balanced training data.
- Trained and compared multiple ML classification models, applying threshold tuning to optimize the precision-recall tradeoff.
- Evaluated model performance using precision, recall, F1-score, and ROC-AUC, visualized via an interactive Streamlit dashboard.

EDUCATION

B.S. Data Science — NED University of Engineering & Technology *Sep 2022 – Present*

- CGPA: 3.86/4.00
- Relevant Coursework: Machine Learning, Database Systems, Algorithms, Data Visualization

Intermediate (Pre-Engineering) — Adamjee Government Science College *Aug 2020 – Aug 2022*

Matriculation — Usman Public School System *Apr 2018 – Apr 2020*

CERTIFICATIONS

Machine Learning Scientist in Python (DataCamp) | Data Analyst in SQL (DataCamp) | Python for Data Science, AI & Development (Coursera) | Power BI Fundamentals (DataCamp) | Huawei Certified ICT Associate – HCIA (Datacom)

Languages: English (Professional), Urdu (Native/Bilingual), German (Elementary), Japanese (Elementary)