MUHAMMAD FAHAD

Data Scientist & ML Engineer
fahadhmughal@gmail.com | +92 302 726 9770 | Lahore, Pakistan
http://www.linkedin.com/in/muhammad-fahad-68580b299 | https://github.com/fahadhmughal

PROFESSIONAL SUMMARY

Software Engineering student specializing in Data Science and Machine Learning, with hands-on Python experience in classification, regression, and predictive modeling. Proficient in scikit-learn, pandas, NumPy, and data visualization tools. Developed and deployed multiple end-to-end ML projects across healthcare and NLP domains. Strong foundation in OOP, DSA, database systems, and cross-functional project collaboration.

EDUCATION

B.S. Software Engineering

COMSATS University
2023 – Present | Lahore | GPA: 3.66 / 4.0

FSC Pre-Engineering

Punjab College of Commerce
2021 – 2023 | Gujranwala | 86.9%

Matric (Science)

The Educators Civil Lines Campus
2019 – 2021 | Gujranwala | 100%

SKILLS

Programming & Querying: Python, Java, C/C++, SQL
Frameworks & Tools: FastAPI, React, Docker, AWS, Azure, Git, GitHub, PostgreSQL
Core Competencies: Data Science, Machine Learning, Artificial Intelligence

PROJECTS

AI-Based Multimodal SAT Tutor Agent (FYP - In Progress)

Designing an AI-powered SAT tutoring system using Llama 3 8B Instruct via Groq API with a RAG pipeline for accurate responses. System supports 4 input modalities: text, image (Tesseract OCR + OpenCV), PDF(PyMuPDF), and voice (OpenAI Whisper). Building React frontend with 3 core screens (chat UI, quiz, progress dashboard) and an adaptive quiz engine supporting 4 SAT subject domains. Backend developed with FastAPI, PostgreSQL, and JWT authentication; deployed on Vercel and Render.

Technologies: FastAPI, LLM, RAG, FAISS, JWT, Groq API, OpenAI Whisper

Medical Insurance Cost Prediction

Built and benchmarked 3 regression models (Linear Regression, Decision Tree, Random Forest) on 1,300+ patient records across 6 clinical features, improving prediction R^2 score by 15% through hyperparameter tuning. Performed EDA, data preprocessing, and feature engineering on structured healthcare dataset. Optimized hyperparameters to improve R^2 score by 15%.

Technologies: pandas, NumPy, Matplotlib, Seaborn, Linear Regression, Decision Tree, Random Forest

Diabetes Prediction using SVM

Trained SVM classifier on 768 patient records from the Pima Indians Diabetes Dataset. Preprocessed 8 medical features including glucose, BMI, and insulin levels for improved model performance. Achieved 77% classification accuracy on held-out test data.

Technologies: Python, scikit-learn, SVM, pandas, NumPy, Matplotlib, SVM

CERTIFICATIONS

Introduction to Ethical Hacking | ***Great Learning*** — **Aug 2024**
Network Security | ***Great Learning*** — **Jul 2024**
Basics of Computer Networking | ***Great Learning*** — **Jul 2024**