Talha Khalid

Lahore, Pakistan | +92-310-7708080 | talhakhalid142@gmail.com
LinkedIn | GitHub | Kaggle

Summary

Aspiring data analyst with hands-on experience in Power BI, SQL, MySQL, and machine learning. Recently completed a degree in Data Analytics and gained practical skills through an internship at PEL, where I worked on Power BI dashboards, SQL queries, and business reporting. Passionate about applying NLP and deep learning techniques in real-world problem-solving, including experience building RAG systems with LangChain and HuggingFace.

Education

BS Data Analytics, Government College University, Faisalabad   2021 – 2025
Final Year Project: Developed a web application for Air Quality Index Prediction using machine learning techniques.

Experience

Data Analytics Intern, PEL Lahore   July 2024 – August 2024

- Performed data analysis using SQL and created interactive Power BI dashboards for sales and collection insights.
- Built and managed a MySQL database to store and structure sales data efficiently.
- Designed DAX-based measures to analyze trends, payment vs. actual collection, and overdue patterns.
- Conducted data cleaning to ensure accuracy and reliability for reporting.

Projects

Air Quality Index Prediction Web App — Python, Scikit-learn, Streamlit   2025

- Built a web application that predicts AQI using machine learning algorithms.
- Cleaned and visualized air quality datasets and integrated the model into a user-friendly interface.

Mental Health Text Classification with LSTM — Python, Keras   2025

- Developed a deep learning model to classify text into: Depression, Anxiety, Stress, or Normal.
- Applied text preprocessing: lowercasing, punctuation removal, tokenization.
- Built a model using embedding layer, Bi-LSTM, dropout, and softmax output.

Medical RAG Question Answering System — LangChain, HuggingFace, FAISS   2025

- Built a Retrieval-Augmented Generation (RAG) system to answer medical questions using custom PDF files.
- Used HuggingFace embeddings, LangChain chains, and FAISS vector store for document retrieval and response.

Skills

- **Programming & Data Science:** Python, SQL, MySQL, Excel
- **Machine Learning:** Scikit-learn, Keras, LSTM, Pandas, Matplotlib

- **Data Visualization:** Power BI, Streamlit, DAX
- **Natural Language Processing:** NLP, RAG Systems, LangChain

Certifications

Learning Data Analytics: Foundations — LinkedIn Learning, 2024

Languages

English (Fluent), Urdu (Native)