Muhammad Dawood Imran

Software Engineer — Data Scientist

+92-3060446517 | portfolio/dawood | linkedin.com/in/dawood | github.com/dawood

SUMMARY

Ambitious and detail-oriented software engineering student with a strong foundation in machine learning, statistics, data analysis, and web technologies. Skilled in Python, data visualization, and automation. Proficient in utilizing Power BI, SQL, and Excel for data processing and visualization. Eager to contribute to innovative AI and ML projects in a dynamic and growth-oriented organization.

EDUCATION

COMSATS Institute of Information Technology   Lahore
Bachelor of Science in Software Engineering (CGPA: 3.50/4.00)   *Sep. 2021 – Present*

Punjab College   Faisalabad
Intermediate (Percentage: 90%)   *Aug. 2019 – May 2021*

TECHNICAL SKILLS

Languages: Python, SQL, R, C, C++

Data Analysis: Data cleaning, processing, visualization, extracting insights, and identifying trends

AI : Machine Learning (Pandas, Scikit-learn, Matplotlib, Seaborn), NLP

Developer Tools: Git, VS Code, Jupyter Notebook, Google Colab

Data Visualization: Excel, Power BI

PROJECTS

Movie Recommendation System | *Python, Pandas, Streamlit, scikit.learn,*
- Created a content-based machine learning model using cosine similarity to recommend movies based on genre, directors, and actors.
- Utilized natural language processing techniques to parse and vectorize movie descriptions and metadata.
- Created a web application using streamlit framework, allowing users to input their preferences and receive movie suggestions.

Heart Attack Risk Prediction | *Python, Pandas, scikit.learn*
- Developed a complete pipeline for heart attack risk prediction, covering data ingestion, validation, transformation, and model evaluation.
- Applied hyperparameter tuning and SMOTE to improve prediction accuracy and handle imbalanced datasets.
- Implemented logging for monitoring and used YAML for configuration management, improving maintainability.

InquireAI - PDF Question Answering Chatbot | *Python, Streamlit, FastAPI, Langchain, ChromaDB*
- An intelligent question-answering system that utilizes Retrieval Augmented Generation (RAG) to provide comprehensive and context-aware responses based on a knowledge base of uploaded documents.
- This project leverages FastAPI for the API, ChromaDB for the database, and Streamlit for the frontend, enabling users to seamlessly interact with the system, upload documents, ask questions, and receive accurate answers powered by advanced AI.

Pakistan Super League (PSL) Batting Analysis | *Python, selenium, Matplotlib, seaborn*
- Scraped data from the ESPN website to identify top players via batting averages, strike rates, and centuries, determined the team with the highest average runs per player, investigated correlations between innings, runs, and strike rates, and visualized findings using bar plots, scatter plots, and histograms.

COURSES

Certified Data Analyst Professional
Completed an industry-recognized certification focused on practical applications of Data Analysis from Digi Skills.

Generative AI: Elevate Your Data Science Career
Learned cutting-edge AI techniques to enhance data science skills, covering generative models and practical applications.