MUHAMMAD UMAR JAN
Data Scientist
Contact: +92 3179919987
GitHub | LinkedIn | Kaggle

Email: umar2exe@gmail.com
Location: Charsadda, Pakistan

EDUCATION

University of Engineering & Technology — Peshawar, Pakistan
Bachelor's Degree in Computer Systems Engineering — 2021 - 2025
Islamia College Peshawar — Peshawar, Pakistan
Pre-Engineering — 2019 - 2021

SKILLS SUMMARY

Languages: Python, C++, SQL
Frameworks: TensorFlow, PyTorch, Scikit-Learn, Keras, Hugging Face Transformers, Hadoop (parallel processing)
Platforms and Tools: VS Code, Google Colab, Jupyter Notebook, Kaggle Notebook, PyCharm, Anaconda, Docker
Areas of Expertise: ML workflows, NLP, LLMs, Lang-Chain, OpenAI, fine-tuning, RAG
Data Handling & Analysis: Feature engineering, data visualization, statistical modeling
Cloud: AWS (sage-maker studio, S3-bucket, EC2, Lambda)

WORK EXPERIENCE

Ezitech Institute — Rawalpindi, Pakistan
Machine Learning Intern — June 2024 – August 2024

- Developed predictive models using Random Forest, XGBoost and ANN on different datasets achieving 93%+ accuracy.
- Collaborated with cross-functional teams to translate business needs into actionable data solutions.
- Developed data models and insights that directly influence strategic decisions made by senior leadership.

National Centre of Big Data and Cloud — Peshawar, Pakistan
Natural Language Processing Intern — February 2025 – Present

- Achieved 95%+ accuracy in NLP tasks using Transformer and TensorFlow.
- Built NLP Pipelines with 80% faster pre-processing using NLTK, Spacy and feature extraction technique like TF-IDF and Word2vec.
- Deployed NLP models on AWS SageMaker, handling end-to-end model training, deployment, and monitoring.

ILLUSTRATIVE PROJECTS

AI-Powered Health Diagnosis & Recommendation Platform(FYP) | GitHub-LINK — 2025

- Create an AI health diagnose platform with Flask and Scikit-learn. Which diagnose symptoms and predict disease.
- Achieved 90.58% accuracy using a Random forest model for condition prediction.

Movie Recommendation App | GitHub-LINK — 2024

- Built a content-based movie recommendation app using NLP Techniques and NLTK to process 5000 movie description.
- Predicted similar movies based on plot summaries and content feature using test similarity technique.

Medical Chatbot Using RAG Technique(FYP) | GitHub-LINK — 2025

- Developed a medical chatbot using RAG with FAISS for vector storage and Mistral-7B as LLM backend.
- Enabled accurate, context aware responses by retrieving relevant medical information from embedded document.

Conversational AI Agent with Multi-Modal Tool Integration | GitHub-Link — 2025

- Developed a voice-interactive AI agent using ElevenLabs for real time speech I/O.
- Implement dynamic tool-driven task executions like web-search, image generation and file operation.

CERTIFICATIONS

- **Python** | Coursera | Certificate
  Completed python course from Michigan covering fundamentals and data structures.
- **Machine Learning** | Coursera | Certificate
  Completed a Machine Learning specialization focused on algorithms and model development.
- **Deep Learning** | Coursera | Certificate
  Completed an advanced Deep Learning program emphasizing neural architectures and practical implementation.
- **Large Language Model** | Coursera | Certificate
  Completed an intensive Large Language Models course covering architecture and NLP applications.
- **Hadoop** | Coursera | Certificate
  Completed a comprehensive course on Hadoop ecosystem and big data processing.
- **AWS** | Coursera | Certificate
  Completed AWS Cloud Essentials course covering core cloud concepts and services.