Muhammad Umer Mansoor

+92 301 5691704 | techsavvyumer@gmail.com | LinkedIn: techsavvyumer | Website: techsavvyumer.pages.dev

WORK EXPERIENCE

Machine Learning 1 | **Data Scientist** | Mar 2024– Present

- Data Scientist with 8+ months of experience in Computer Vision, and 2+ months of experience in OCR (Optical Character Recognition).
- Designed and implemented data-driven solutions deployed in production environments, serving millions of users.
- Contributed to the development of high-impact features for a top 5 company in its domain.
- Hands-on experience in Retrieval-Augmented Generation (RAG) and building Knowledge Graphs to enhance product functionality.
- Developed and implemented a full-fledged Attendance System using facial recognition with CCTV cameras, handling both frontend (Streamlit) and backend (FastAPI) development entirely on my own. The system is now actively used within the company for attendance tracking.
- Designed and implemented machine learning models for tasks including facial recognition, classification, and price prediction, leveraging libraries like Python and frameworks like TensorFlow.
- Conducted extensive data mining and scraping techniques using Selenium, Beautiful Soup, and hidden APIs to extract valuable data from diverse sources for training the model.
- Generated synthetic datasets to enhance training and testing of deepfakes detection models, addressing data scarcity, and improving model robustness.

Knowledge Streams | **DS & ML Trainee** | Oct 2023– Jan 2024

- Learned exploratory data analysis, statistical analysis, data visualization, and machine learning algorithms, developing strong analytical thinking and research skills.
- Managed time effectively to complete 3 major projects while meeting tight deadlines, demonstrating strong project management abilities.
- Collaborated cross-functionally to implement machine learning solutions, exercising teamwork and communication skills.

PROJECTS

Automatic License Plate Detection and Recognition | **Best Project @ Knowledge Streams** | YOLOv8, OpenCV, CNN

Designed a vehicle and license plate detection system using YOLOv8, created an OpenCV pipeline for character segmentation, and developed a CNN model for license plate recognition, achieving higher accuracy than traditional methods.

Duplicate Bug Report Detection and Classification | **FYP @ VU** | NLTK, Spacy, Scikit-learn

Preprocessed bug report data through tokenization, stopword removal, and lemmatization, built a classifier using Naive Bayes, SVM, and Random Forest to identify duplicate bug reports with 82% precision, and recommended model enhancements by analyzing misclassifications to improve developer efficiency.

Realtime Tweets Sentiment Analysis | **Winner Project @ Datathon Competition** | Tweepy, NLTK, Streamlit

Developed a live Twitter sentiment analyzer app using Streamlit, fetching data via the Twitter API, achieving 75% accuracy by cleaning and preprocessing tweets with regex and NLTK, and streamlined the data workflow from tweet acquisition to sentiment prediction, showcasing strong programming and project management skills.

CORE SKILLS

Programming Languages: Python
Developer Tools: VS Code, Jupyter Notebook, Git, Conda, MLflow, Linux, Bash
Frameworks: Scikit-learn, TensorFlow, Keras, PyTorch, OpenVINO, HuggingFace
Web Development Frameworks: Streamlit, FastAPI
NLP & Retrieval-Augmented Generation (RAG) Frameworks: NLTK, LangChain, LlamaIndex
Databases: MySQL, Microsoft SQL Server
Cloud Computing: Azure, AWS, GCP

CERTIFICATIONS

Google Advanced Data Analytics | Coursera | 2023

EDUCATION

Virtual University of Pakistan, Lahore
BS Computer Science, Artificial Intelligence | 2019 – 2023