

Junaid Ali

📍 Lahore, Punjab, Pakistan ✉ junaida.rahmat@gmail.com ☎ +923034142029 🔗 in/junaid-ali-10021a14a 🌐 github.com/JunaidARahat

SUMMARY

A dedicated and inspiring Data Scientist with an objective of working in organization that provides opportunities for technical and personal advancement.

PROJECTS

US Visa Approval Prediction | github.com/JunaidARahat/US-Visa-Approval-Prediction/

- Objective:
- Developed and deployed a machine learning model to predict the approval of US visa applications, incorporating MLOps practices for scalability and reliability.
- **Tech Stack:** Python, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, MongoDB, Fast API, Docker, AWS CICD Deployment with Github Actions
- **Key Contributions:**
 - Conducted data preprocessing, feature engineering, and exploratory data analysis to identify key predictors.
 - Developed and evaluated various classification models, including logistic regression, random forest, and XGBoost, achieving an accuracy with K-Nearest Neighbour of [96]% on the test dataset.
 - Integrated MLOps practices by containerizing the model with Docker, automating deployment AWS CI/CD Deployment with Github Actions.
 - Implemented monitoring and logging to track model performance and ensure continuous integration and deployment.

Chest Disease Classification | github.com/JunaidARahat/Chest-Disease-Classification-from-Chest-CT-Scan-Image/

- **Objective:** Developed a deep learning model for classifying chest diseases using CT scan images, with an emphasis on reproducibility and deployment.
- **Tech Stack:** Python, TensorFlow, MLflow, DVC, Flask, Docker
- **Key Contributions:**
 - Preprocessed and augmented CT scan images to enhance model robustness and performance.
 - Trained a convolutional neural network (CNN) for disease classification, achieving [90]% accuracy on the validation dataset.
 - Employed MLflow for experiment tracking, versioning, and model management to ensure reproducibility.
 - Utilized DVC for data version control, enabling efficient collaboration and tracking of dataset changes.
 - Deployed the model as a Flask web application, containerized with Docker for easy deployment and scaling.

Named Entity Recognition (NER) Project | github.com/JunaidARahat/Name-Entity-Recognition

-
- **Objective:** Developed a Named Entity Recognition model using BERT for identifying and classifying entities in text, with deployment on Google Cloud Platform (GCP).
- **Tech Stack:** Python, BERT, Hugging Face Transformers, FastAPI, Google Cloud Platform (GCP), CircleCI, Docker.
- **Key Contributions:**
 - Fine-tuned the BERT model for NER tasks, achieving [95]% F1 score on the validation dataset.
 - Developed a RESTful API using FastAPI to serve the NER model for real-time predictions.
 - Deployed the API on GCP, ensuring scalability and high availability.
 - Implemented continuous integration and deployment (CI/CD) using CircleCI, enabling automated testing, building, and deployment.
 - Containerized the application with Docker for consistent and portable deployment across environments.

Finance Chatbot | github.com/JunaidARahat/Finance-Chatbot

- **Objective:** Developed a finance-focused chatbot capable of providing insights and recommendations on financial queries, utilizing Open AI, Langchain and astradb for advanced natural language processing.
- **Tech Stack:** Python, Langchain, AstraDB, Flask.
- Built and integrated Langchain for natural language processing, enabling the chatbot to understand and respond to complex financial queries.
- Utilized Astradb for scalable and efficient data storage, ensuring quick access to financial data and insights.

- Developed the chatbot as a Flask web application, providing a user-friendly interface for real-time interactions.
- Enhanced the chatbot's ability to deliver accurate financial recommendations, improving user engagement and satisfaction.

Research Paper Summarization Using Generative AI | github.com/JunaidARahat/Research-Paper-Summarization-Using-Generative-AI

- **Objective:** Developed a system for automatically summarizing research papers, enabling users to quickly grasp key insights and findings using advanced NLP techniques.
- **Tech Stack:** Python, LangChain, ChromaDB, OpenAI, Streamlit, Gradio.
- Integrated LangChain with OpenAI's language models to generate concise and accurate summaries of complex research papers.
- Utilized ChromaDB for efficient data storage and retrieval, allowing for rapid processing of large datasets.
- Developed an interactive web interface using streamlit and gradio, enabling users to upload and summarize research papers in real-time.
- Enhanced the system's summarization accuracy through iterative testing and fine-tuning of the underlying models.

EDUCATION

Master of Information Technology | Virtual University | Lahore, Pakistan | 2017 | 3.21

Bachelor of Arts | Punjab Univerity | Lahore, Pakitan | 2011

Fsc Pre Engineering | Dayal Singh College | Lahore, Pakistan | 2009

CERTIFICATIONS

MLOPS: Production Ready Data Science Projects | iNeuron | <https://learn.ineuron.ai/certificate/260013ae-3e00-4bd9-8afc-931166c28c0c>

SKILLS

Python

Machine Learning (Scikit-learn | Pandas | NumPy)

Deep Learning (Keras | TensorFlow | Pytorch)

Natural Language Processing (Bert | LSTM | RNN | GPT)

Large Language Models (Hugging Face | Open AI | Open Source LLMs)

REST API (Flask | Fast API)

MLOps (GIT | DVC | Docker | GitHub Actions | Circle CI)

Frameworks (LlamaIndex | LangChain)

Cloud Platforms (AWS (ECR | EC2 | S3 |) | GCP(VM Instance |Artifact Registry | Cloud Storage) | AZURE)

Database: MySQL, MongoDB