

Anas Khan - Data Scientist

Contact Information Phone No: 8329796747 Email: k.anas4982@outlook.com [LinkedIn](#) [GitHub](#) [Kaggle](#) [LeetCode](#)

Associate Data Scientist with 3+ years of experience in building and deploying machine learning and DL models.

Technical Skills:

Programming Languages: Python, CPP, JAVA, SQL, NoSQL, JavaScript, Flask

Database Management: PostgreSQL, Casandra, MySQL, MS SQL Server, MongoDB

Machine learning & NLP: OpenCV, Deep Learning, Scikit-learn, Sci-Py, TensorFlow, PyTorch, Pandas, NumPy

MLOps & Deployment: MLFLOW, Airflow, DagsHub, Streamlit, GitHub, AWS, Azure

Data Analysis & Visualization: Pandas, Power BI, Excel, Matplotlib, Seaborn

Work Experience:

Smart Software Services (I) Pvt. Ltd | Associate Data Scientist | Full time | Mar 2022 – Present

- Designed and deployed predictive models using Python, Scikit-learn, and TensorFlow, improved model accuracy.
- Analyzed HR data for insights on attrition data for a confidential fortune 500 Co. using traditional data science, ML, NLP techniques, LLM embeddings and text generation for classification.
- Collected, cleaned, and preprocessed data to develop a production-grade Machine Learning recommendation engine leveraging CLIP and vector search.
- Implemented MLOps techniques to streamline deployment, monitoring, and scalability of the model.
- Conducted exploratory data analysis (EDA) using Pandas and Matplotlib, uncovering key sales trends that boosted revenue strategies.
- Cleaned, pre-processed, and validated data using SQL and Python, reducing data processing time by 35% through optimized queries.

Projects:

AI-Powered Recommendation System

- Designed and deployed an AI-driven recommendation system leveraging CLIP for image and text-based retrieval.
- Implemented **vector search** using **FAISS** to enhance similarity-based search efficiency.
- Developed a **scalable API backend** with **FastAPI** for seamless integration and performance.
- Deployed the system using **Docker** and **Kubernetes** on **AWS**, ensuring scalability and reliability.
- Applied MLOps best practices for model monitoring, logging, and automated retraining.
- Optimized **embedding search performance**, reducing retrieval time and improving recommendation accuracy.
- Integrated the system with a **user-friendly frontend**, enabling seamless interaction and personalized recommendations.

Consumer Complaint Resolution Prediction

- Developed a **predictive model** to determine whether a consumer complaint resolution would be disputed.
- Conducted **exploratory data analysis (EDA)** to identify patterns and improve feature selection.
- Handled **missing values**, **encoded categorical variables**, and extracted text features using **TF-IDF**.
- Built and fine-tuned **classification models** using machine learning algorithms to enhance prediction accuracy.
- Evaluated model performance with precision, recall, F1-score, and ROC-AUC to ensure reliability.
- Improved prediction accuracy, enabling early intervention and reducing dispute rates by 30%.

Education

Master of Computer Application

Oct, 2021 – May, 2023

M.G.M's Jawaharlal Nehru Engineering College Aurangabad MH, India | CGPA: 8.64 | Awarded Gold Medal

Bachelors in Computer Science

May, 2018 – Sep, 2021

Dr. Babasaheb Ambedkar Marathwada University Aurangabad MH, India | Percentage: 74

Accomplishments

Winner- Won Chancellor Gold Medal among all of the PG programs in the faculty of Engineering and Technology, MGM University for the academic year 2022-23