Hammad Anjum
Data Scientist

+92-3237734497 | hammad95632@gmail.com | Lahore, Pakistan
LinkedIn/hammad6271 – Portfolio

SUMMARY

Data Science Senior with a robust foundation in machine learning, Artificial Intelligence, and Natural Language Processing. Bringing hands-on experience from academic projects in predictive modeling, data visualization, and exploratory data analysis. Proficient in Python, R, SQL, and expert in machine learning algorithms and AI techniques, along with data analysis, preprocessing and model building/evaluation.

PROJECTS

PDF Text Retrieval Augmented Generation
- Utilized Streamlit and OpenAI's advanced text embedding model to develop a web application for PDF text retrieval, significantly enhancing user access to document content by 5%.
- Leveraged ChromaDB to optimize the storage and retrieval of document embeddings, resulting in faster access and improved data management by 10%.
- Implemented user-friendly functionalities in the web application allowing for PDF uploads and precise query entries, which streamlined the information retrieval process for users.

YouTube Comments Classifier
- Developed a web scraping script using Selenium and Beautiful Soup, to collect over 1000 comments from prominent YouTube channels.
- Added a robust cleaning pipeline in Python, including spell checking, lemmatization, and language filtering, resulting in a refined dataset with a 20% reduction in noise.
- Utilized machine learning models such as Logistic Regression and Naive Bayes, achieving an accuracy of 70% and 67%.
- Implemented deep learning models, including Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), achieving competitive accuracy rates of 45% and 52%.

Arline Satisfaction Analysis
- Conducted comprehensive data preprocessing on a dataset with over 100,000 entries and 20+ features, including handling missing values and removing duplicates.
- Implemented and evaluated 7 machine learning models, achieving up to 90% accuracy in predicting airline customer satisfaction.
- Applied techniques such as RandomOverSampling to balance the dataset and GridSearchCV to optimize model hyperparameters, enhancing model performance by 5%.
- Developed detailed visualizations and performed exploratory data analysis to uncover key insights and data patterns.

SKILLS

Data Analysis & Visualization: Power BI (Advanced), Tableau, SQL, python libraries (Seaborn, Matplotlib, Plotly)
Programming & Databases: Python (Pandas, NumPy, SciPy), R, SQL, MSSQL
Data Science & Machine Learning: RapidMiner, Weka, TensorFlow, Keras, Jupyter, PyTorch, NLTK, openAI, HuggingFace, OpenCV
Cloud Technologies: Google Cloud Platform (GCP: Dataproc, Dataflow, BigQuery)
Soft Skills: Communication, Leadership, Time management, Dedication, Hard working
Techncical Skills: Data preprocessing, Data manipulation, Model building, Model evaluation, Natural language processing, Data visualization, Data analysis, Data analytics, Exploratory data analysis, Computer vision.

EDUCATION

FAST National University (NUCES) – Lahore, Pakistan   **June 2025**
Bachelor of Science – Data Science
- Concentration in Machine Learning and Artificial Intelligence
- Cumulative GPA: 3.18/4.0

CERTIFICATIONS AND AWARDS

- **Professional Data Scientist** - **DataCamp**   **July 2024 - July 2026**
- **Data Analyst Associate** - **DataCamp**   **October 2023 - October 2025**

- [Data Analyst with Power BI – Track](#) - Datacamp   **March 2024**
- [Google Cloud Skills Badges](#) - Google   **May 2024**