RAFAY **MAHMOOD**
Data Scientist - III

rafaymahmood1999@gmail.com
+92 331 4336 177
linkedin/rafayaar | github/rafayaar

Skills

Machine Learning & AI
- Large Language Model (LLM)
- Knowledge Graphs
- Vector Databases
- RAG Engines
- LLM Agents, Functions & tools
- Semantic Similarity Search
- NLP – NER – Transformers
- Hyperparameter Tunning
- CNN - RNN - Fine Tuning
- TensorFlow - Keras - PyTorch

Data Science & Engineering
- Statistical Analysis – ETL
- Indexing
- EDA - Apache Spark
- SQL - NoSQL - DBMS

Deployment & Operations
- Hugging Face - Inference APIs
- AWS - Azure
- GitHub - Git - Flask - Docker

Education

University of Central Punjab
BS Computer Sciences
2018 - 2022 | CGPA: 3.17 (A)

Punjab Group of Colleges
F.S.c - Pre-Engineering
2016 - 2018 | Grade: A

Coursework

Object Oriented Programming
Data Structures
Machine Learning
Deep Learning
Data Science
Artificial Intelligence
Big Data Analytics
Database Administrations
Probability and Statistics

Extras

Video Games
Gymnasium, Badminton

Professional Experience

Foretheta Group May 2023 – Current
Data Scientist III
- Led the improvement of a data pipeline managing 95 million records, utilizing MP-NET to achieve a 2x accuracy boost over the previous system.
- Used advanced models such as GPT-NeoX, BERT,MP-NET ST to enhance the patent data index, improving prior art search
- Created a robust analytics tool with OpenAI API for extracting insights from extensive patent datasets.

Reacon Group May 2023 – Current
Machine Learning Engineer
- Leveraging OpenAI agents, functions, and tools, I actively contributed to the development of a Conversational AI chatbot product using LangChain and GPT-4.
- Empowered by LangChain and GPT-4 with OpenAI tools, ensured over 90% accuracy of the final MVP

Punch Group LLC Dec 2022 – May 2023
Machine Learning Engineer
- Trained and Deployed on server, a classification model to identify Sales Email prospects having accuracy of more than 90%.
- Working on an open source LLM named BLOOM (*Big Science*) to auto generate responses to emails with an accuracy over 75% compared to GPT-3. Exploring other LLMs such as Flan-T5, GPT.

Leverify LLC Oct 2022 – Dec 2022
Business Automation Executive
- Engaged in cross-functional collaboration to automate the supplier MPN matching process for Amazon listing products, resulting in a system that achieves a 73% accuracy rate. Enhanced the prior solution, doubling its processing speed while significantly boosting accuracy.

Radical X Aug 2022 – Oct 2022
Virtual Project Manager Intern - Python
- Led a project focused on developing an unsupervised learning algorithm to form intern teams based on various factors, including geographical regions, time zones, and countries

Upwork - Freelance Jan 2021 – Dec 2021
Machine Learning Engineer
- Worked on projects related to data driven fields such as ML, DL, DS, NLP, OCR etc with the job success of more than 90%.

SoloInsight Inc. Aug 2019 – Sep 2019
PHP Development Intern
- Used *PHP* and *SQL* to manage how data is stored in tables and user interact with it.

Projects

Squad Matching Algorithm Aug 2022 – Sep 2022

- Created a system to assist in forming intern teams, emphasizing the alignment of team members based on shared characteristics while prioritizing individual comfort. Employed K-Means clustering to achieve this objective, utilizing Python and essential libraries like Pandas and Matplotlib to present data insights effectively.

NLP Text Extraction Jul 2022 – Aug 2022

- Developed a script capable of extracting text from both images and PDF files, followed by data cleansing, parsing, and tokenization to align with specific requirements, mirroring the concept of resume parsing commonly employed by companies. This project was executed using Python, along with libraries such as Tesseract, PyPDF, and NLTK

AI based fitness app using CNN Oct 2021 – Mar 2022

- Utilized the CNN technique and primary tools, including *TensorFlow*, *Keras*, and *Python*, to train a Machine Learning model for an Image Classification System, achieving an accuracy rate exceeding 75%.

Appo-Fit Application Mar 2022 – July 2022

- Developed a *Flutter* application seamlessly integrated with an Image Classification model, enabling predictive capabilities and leveraging *Firebase Firestore* as the underlying database.

Heart Attack Analysis July 2021 – July 2021

- Conducted an analysis of cardiac illness patient data to identify the primary factors (vitals) contributing to heart disease, utilizing *Python* and data analytics tools, including *Seaborn*, *Matplot*, *Pandas* and *NumPy.*

Leadership, Awards and Certs.

ML Research Paper submitted to ISIEA2023 March 2023

- The investigation focused on food images, aiming to explore how various CNN models respond to different data inputs.

Director Logistics Mar 2022 – July 2022

- Served as a Director of Logistics on a voluntary basis within a student-led campus organization.

National Winner at Global Battle of Minds'21 by BAT Nov 2021 – Dec 2022

- Participated in the Global Finals, representing Pakistan among 13 countries on the international stage.

Samsung I. Campus Oct 2021 – Feb 2022

- Successfully obtained admission to an AI training course offered by Samsung.

Python by University of Michigan
Python for Data Science, AI and Development by IBM
Project Data Science by IBM
Machine Learning by Stanford (Andrew Ng)