

RAFAY MAHMOOD

[in LinkedIn](#) | [+92 331 4336177](#) | [rafayaarm@gmail.com](#) | [GitHub](#)

Skills

- **Data Science & Engineering:** Python | R | MySQL | Predictive Analysis | Statistical Modelling | Pipelines | EDA | Flask | Django | Pre / Post Processing | Visualization | Spark | Hive | ETL | ELT | Big Data | Software Engineering | CI/CD
- **AI / ML:** Generative AI | NLP | NER | Recommendation System | Large Language Model (LLM) | Vector Databases | LangChain | Agents | Chatbots | Finetuning | RAG Engines | Semantic Similarity Search | CNNs | RNNs | LSTMs | Transformers | TensorFlow | PyTorch
- **Cloud & Operations:** AWS | GCP | Version Control | Git | GitHub | Docker | Inferences | APIs | Linux | Bash

Experience

Data Scientist - III Foretheta *Seattle, USA - Remote* **05/2023 - Current**

- Led the improvement of a complex data pipeline managing 95 million records, employing the advanced sentence transformer model MP-NET to significantly enhance processing capabilities, resulting in a remarkable 2x accuracy boost over the previous system.
- Used batch transform jobs on AWS SageMaker for data vectorization, followed by creating a FAISS index as a vector database. The indexing process was performed on multiple GPU based EC2 instances. Later, Flask was used along with AWS API Gateway and CloudWatch integrations to ensure smooth operations. S3, Amplify, Lambda and other supporting services were also used.
- Created a robust analytics tool with the OpenAI API for extracting insights from extensive patent datasets, enhancing the capability to derive actionable intelligence.
- Integrated Continuous Integration/Deployment pipelines, managed pull requests, conducted code reviews, and performed load/stress testing, as well as unit, integration, and end-to-end testing
- Collaborated with cross-functional teams to optimize performance and scalability, leveraging advanced machine learning models and cloud infrastructure to meet evolving business needs.

AI Engineer Reacon Group **05/2023 - Current**

- Developed an AI-powered feature that significantly improved the quality and effectiveness of marketing content across various channels, including direct mail, email, digital advertisements, and SMS. This innovation enabled users to craft more compelling and personalized messages, leading to higher engagement rates and better campaign performance.
- Used Llama2 to cut the time needed to create marketing campaigns by over 60%, enabling faster and more frequent launches.
- The integration of these AI features resulted in measurable improvements in marketing productivity and effectiveness. The reduction in campaign creation time by over 60% and the ability to quickly analyze large datasets through the chatbot contributed to faster turnaround times, higher engagement rates, and more successful marketing initiatives.

Machine Learning Engineer Punch Cool **10/2022 - 05/2023**

- Trained and deployed on production, a classification model to identify Sales Email prospects having accuracy of more than 90%.
- Actively researching and evaluating effective solutions using Large Language Models (LLMs) like BLOOM to address various AI tasks, aiming to enhance the functionality and performance of AI applications
- Developing robust classification models to accurately differentiate and categorize various types of emails. This effort improves the organization and handling of email communications, streamlining workflow and response management.
- Implementing multi-label classification techniques to identify and categorize multiple intents within email content. This automation significantly speeds up the product development process by ensuring accurate intent recognition and handling

Apprenticeship – Python Radical X **08/2022 - 10/2022**

- Led an intern project focused on developing an unsupervised learning algorithm to form intern teams based on various factors, including geographical regions, time zones, and countries.

Software Engineer, Intern SoloInsight **08/2019 - 09/2019**

- Was a part of database team where I had to pre write queries for Senior SE's to actually review and process them for dev or prod use.

Education

Bachelor's of Computer Science University of Central Punjab *CGPA: 3.17(A)* **2018 - 2022**

Projects

- **APPOFIT:** Created an AI powered Flutter Application for Gym trainers to keep track on trainees by improving communication **(03/2022)**
- **RED:** Idea of Restoring Environment from Depletion was to reduce carbon footprint and improve process of recycling waste **(11/2021)**

Others

- **Research Paper:** Paper published related to Computer Vision on Food Images using CNN at ISIEA2023 conference. [Paper Link](#) **(03/2023)**
- **National Winner Global Battle of Minds:** As a National Finalist represented Pakistan in global finals among 13 countries. **(11/2021)**
- **Early Bird AI:** Got selected for Samsung Innovation Campus AI training as an early bird. Covered most of AI curriculum before col courses.