

MOHAMED AFRITH

Data Scientist / AI ML Engineer / Data Analyst

- ✉ mafrith007@gmail.com
- ☎ 9489214174
- 📍 175, North Street, Ammapattinam, Pudukkottai district, TamilNadu
- 🌐 linkedin.com/in/mohamed-afrit-s/
- 🐙 github.com/mhdafriith

WORK EXPERIENCE

AI Data Analyst - Recykal

- 03/2025 – Present | Hyderabad
- Led end-to-end development of AI workflows by processing 10,000+ records using Python and Pandas, training/testing a custom Donut model, creating 4+ performance dashboards, and building a RAG-based chatbot with LangChain, FAISS, and LLaMA3 (via Groq API); leveraged LlamaParse for extracting structured data from 100+ PDFs and optimized LLM prompt design to boost response accuracy by 40% and model evaluation efficiency by 25%.

SKILLS

Gen AI (Langchain, Langsmith, Langraph, Donut)

Python

Agentic RAG

CrewAI

Seaborn

MySQL

Pandas

MLflow(MLOPS)

Looker studio

Streamlit

AWS(EC2, S3)

MS-Excel

NLP(Spacy)

Paddle OCR

Selenium

ML(Scikitlearn)

Vision language Model

Prompt Engineering

n8n

LLM

Agentic AI

DL (Pytorch, Tensorflow)

Vector Database(Qdrant, FAISS, ChromaDB)

ACHIEVEMENTS & AWARDS

- Best Innovation Award at a National Level Hackathon - SRM University.**
- **Developed a machine learning model** to predict business profit from financial and operational data, improving accuracy over baseline approaches. **Built and deployed an interactive Streamlit application** to showcase real-time profit prediction and insights, which won the

PROFILE

- **Data Scientist & AI/ML Engineer** skilled in data preprocessing, model training, LLM prompt engineering, and RAG-based chatbot development, with proficiency in Python, SQL, LangChain, Hugging Face, and AWS. Recognized with the **Best Innovation Award** in a national ML Datathon, holding a **Master's in Data Science (IIT-M GUVI)** and an **MBA (SRM Institute)**.

PROJECTS

- Document Understanding & Data Extraction (GenAI) - Recykal**
- Built an **intelligent document automation system** to streamline classification and structured key-value extraction, using **Donut (Vision Transformer)** for document categorization.
 - Designed a multi-layer **extraction pipeline** with Donut for vision-language understanding, **LlamaParse + LLM** for semantic parsing, and **OCR + LLM** fallback for low-quality scans, followed by a **validation layer** to ensure data accuracy.
 - **Optimized processing workflows** by integrating the solution with existing enterprise systems, enabling smooth adoption and faster decision-making.
 - Scaled the system to process **60K+ documents/month** with **92% accuracy**, achieving a **7x productivity boost**, reducing manual errors, and enabling **real-time compliance-ready insights**.
- Key Skills:** Donut (Vision Transformer), LlamaParse, LLMs (OpenAI, Llama), OCR + LLM, Document Classification, Key-Value Extraction, GenAI Pipelines, Automation, Data Validation, Deep Learning
- Agentic RAG Chatbot for Research Papers & Live News**
- **Developed** an interactive RAG-based chatbot using **Python, Streamlit, and LangChain**, integrating **FAISS** vector search over 100+ arXiv research papers for semantic retrieval and live news search via **CrewAI agents**.
 - **Engineered** embeddings with **HuggingFace all-MiniLM-L6-v2** and integrated **Groq LLaMA3-70B** to deliver high-quality, context-aware answers, reducing research query resolution time by **~60%**.
 - **Built** a hybrid AI pipeline to automatically route non-research queries to a CrewAI-powered **News Analyst Agent**, enabling **real-time information access** and improving user engagement by **~40%** during testing.
- Key Skills:** Python, Streamlit, LangChain, FAISS, HuggingFace Embeddings, Groq API, LLaMA3-70B, CrewAI, RAG Pipeline, Semantic Search, Agentic AI, DuckDuckGo Search
- My Zoom: A Transformer-Based Model for Contextual Feedback Validation**
- **Developed** a Transformer-based feedback validation model using BERT, achieving **91.2% accuracy** and improving contextual feedback classification by **23% over traditional methods**.
 - **Automated** preprocessing and analysis of **15,000+ feedback entries**, reducing manual review effort by **60%** through contextual understanding and classification.
- Key Skills:** Transformers (BERT), NLP, Text Classification, HuggingFace, PyTorch
- Predicting Depression from Mental Health Survey Data using Deep Learning**
- **Developed** a deep learning model using TensorFlow/Keras to predict depression from a dataset of **3,000+ survey responses**, achieving **91% accuracy** and **0.92 AUC-ROC**.
 - **Deployed** the model via a user-friendly Streamlit app, enabling real-time mental health screening with dynamic visual insights and streamlined user interaction.
- Key Skills:** Streamlit, Binary Classification, Cloud Deployment

EDUCATIONAL QUALIFICATIONS

- IIT-M Advanced Programming Professional & Master Data Science, IIT- M GUVI**
07/2024 – 03/2025 | Chennai
- MBA**
SRM Institute of Science and Technology
2023 – 2025 | Chennai