Jeet Majumder

Data Scientist — Machine Learning & AI

+91-6297XXXXXX | jeet0912majumder@gmail.com | jeetmajumder.vercel.app | West Bengal, India

Summary

Data Scientist with hands-on experience in machine learning, NLP, and LLM-based systems. Built end-to-end ML and RAG pipelines with a focus on scalable, production-ready AI solutions. Passionate about translating ideas into reliable, real-world AI systems.

Skills

Programming: Python (Pandas, NumPy, Scikit-learn)
Machine Learning & AI: Machine Learning, Deep Learning, NLP, Computer Vision, LLMs, RAG
ML & AI Tools: TensorFlow, PyTorch, OpenCV, YOLO, LangChain, Ollama, Hugging Face
MLOps & Deployment: MLflow, DVC, Docker, Kubernetes, GitHub Actions (CI/CD), AWS, Streamlit, Render

Experience

Junior Data Scientist **Jul 2025 – Sep 2025**
Databae Technologies LLP *Tech Stack: Computer Vision, Re-identification*

- Built a **real-time person tracking system** using body and face embeddings with re-identification techniques to handle occlusion and overlapping scenarios in CCTV footage.
- Designed a robust **person entry–exit detection system** for door-level monitoring, capturing timestamped images and handling edge cases efficiently, **reducing reliance on external LLM APIs and associated costs**.
- Integrated the tracking and entry–exit detection pipelines with backend services in collaboration with the team.

Data Science Intern **Apr 2025 – Jul 2025**
Databae Technologies LLP *Tech Stack: Machine Learning, Deep Learning, Computer Vision*

- Built and trained a **YOLO-based** object detection model for **spark and person detection in workshop environments** to analyze workers' daily activity patterns.
- Worked on ML and DL model development, fine-tuning, and evaluation during early-stage experimentation.

Projects

RAG-based Document Question Answering System | *Python, NLP, LLMs* | **Nov 2025 – Dec 2025**

- Built a production-ready **Retrieval-Augmented Generation (RAG) system** for document-based question answering using local LLMs.
- Implemented **end-to-end retrieval pipelines** including text ingestion, chunking, embeddings, similarity search, and cross-encoder re-ranking, with optimized context construction using Stuff and Map-Reduce strategies to **reduce hallucination and improve answer reliability**.
- Designed the system to **balance response speed and accuracy** while providing source-backed answers for transparency.

News Topic Classification App | *Python, NLP, ML* | **Oct 2025 – Nov 2025**

- Developed an **end-to-end NLP system** that ingests raw news articles and serves real-time predictions to classify news such as World, Sports, Business, and Technology via a FastAPI service.
- Created modular data preprocessing, training, and inference pipelines following **production-ready MLOps practices, with automated CI and Dockerized deployment.**
- Optimized the inference pipeline for **low-latency classification** while maintaining stable model accuracy in real-time usage.

Awards & Research Contributions

- **International Conference Paper Presentations:** Presented three research papers at international conferences on topics including Drone Data Transmission, Hepatitis C Prediction, and Brain Tumor Detection.
- **Best Final Year Project (2025):** Developed an IoT-based machine learning system for heart disease classification, integrating real-time sensor data with an ML model to generate diagnostic reports via a web application.

Education

B.Tech CSE Specialization in Data Science **Graduated: 2025**
Brainware University, Kolkata *CGPA: 9.56 / 10*