UDDIPAN BASU BIR

Address: StubenlohStr.6,91052Erlangen, Germany
Phone: +49 17667314674
Email: uddipanbb95@gmail.com
LinkedIn: https://www.linkedin.com/in/uddipan-basu-bir/
Portfolio: https://uddipan77.github.io/portfolio_Uddipan/

SUMMARY

- Masters Degree in Data Science with specialization in Artificial Intelligence and Machine Learning
- Bachelors Degree in Computer Science and Engineering
- Experience working as Data Analyst, Data Engineer, AI & ML Engineer

WORK EXPERIENCE

Research Team Member, Munich Music Labs, TUM Germany **Nov 2025 - Present**

Responsibilities :

- Developing a LLM/foundation-model–driven pipeline to transcribe audio into symbolic music .

Working Student – Applied AI Engineer, Siemens AG Germany **May 2025 - Present**

Responsibilities :

- Working on AI use cases on Helix AI framework and RAG pipelines with Azure OpenAI.
- Creating a PoC of Agent-based modelling using AgentTorch and LLM based policy.

Tools & Technologies : Microsoft Co-pilot Studio, Azure Machine Learning, Azure Open AI, LangChain, PyTorch, Azure Cosmos, Azure AI Search, Helix AI, Postman Collection

Working Student - Data Analyst, Schaeffler Germany **June 2024 - April 2025**

Responsibilities :

- Development of interactive Microsoft PowerBI dashboards for BOM and Financial Data.
- Data wrangling, preprocessing and data modeling. Built low-code business apps in Microsoft
- PowerApps, integrating SharePoint and Dataverse and automating workflows with Power Automate.

Tools & Technologies : PowerBI, Power Apps, Power Automate, SAP, Dataverse, Microsoft Power Platform

Tata Consultancy Services, Kolkata, India **Sep 2018 - June 2022**

Data Engineer **Sep 2021 - June 2022**

Responsibilities :

- Re-engineered legacy Ab-initio ETL pipelines into PySpark and Airflow.
- Automated extraction and loading of data from Unix servers and Oracle.

Data Analyst **Sep 2018 - Aug 2021**

Responsibilities:

- Design and maintain efficient data pipelines, dashboards, and analytics solutions to support decision-making.
- Experienced in financial data analysis including MRR, ARR, and P&L reporting.

Tools & Technologies : Kubernetes, Kafka, Airflow, GitHub Enterprise, Python, Unix, Autosys, Abinitio, SQL Indexing, Window Functions, CTEs, SQL Procedures, ServiceNow, PowerBI, PowerApps, SQL, Python

EDUCATION

M.Sc.Data Science, specialization in AI & ML **Oct 2022 - Present**

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

- Courses: Deep Learning – 2.7, Pattern Recognition – 1.3, Advanced topics in Deep Learning – 1, Machine Learning in Signal Processing – 1.7, Machine Learning in MRI – 1.3
- Master Seminar Project (REAHU & FAU): Forecasting and Anomaly detection in Time Series Data for polymer extrusion process. GitHub - https://github.com/uddipan77/REHAU-Digital-Twin/tree/main

Bachelor of Technology in Computer Science & Engineering **June 2014 - July 2018**

Maulana Abul Kalam Azad University of Technology, Kolkata, India

- Bachelor Thesis: https://github.com/uddipan77/Emotion-Detection-Through-Facial-Expressions

TECHNICAL SKILL

- **Programming Languages:** Python, SQL, Unix Shell Scripting, TypeScript
- **Machine Learning & Deep Learning Frameworks:** PyTorch, PyTorch Lightning, Scikit-learn, XGBoost, statsmodels
- **NLP / LLM Tooling:** Hugging Face Transformers, LangChain, LangGraph, Groq Cloud, SWIFT (LLM fine- tuning), LLaMA-Factory, NLTK, spaCy, vLLM, LangFlow, n8n
- **Vector Databases:** OpenSearch, Qdrant, FAISS, Chroma
- **Data Validation & Model Serialization:** Pydantic, ONNX
- **CI/CD & Deployment:** GitHub Actions, Docker, Kubernetes
- **Data Engineering & Orchestration:** PySpark, Apache Spark, Airflow
- **Cloud & High Performance Computing:** Azure, AWS, HPC
- **Experiment Tracking, HPO:** Opik**,** MLflow, Weights&Biases (WandB), Tensorboard, Optuna, ZenML
- **Web & API Frameworks:** FastAPI, Flask, Node.js, Streamlit
- **Data Visualization & BI:** PowerBI, Power Automate, Power Apps
- **Observability & Monitoring**: Prometheus, Grafana

KEY PROJECTS

MASTERTHESIS: From Pixels to Structure: Analysis of Lightweight Vision- Language Models for Document OCR and Structured Output Generation **June 2025 - Present**

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

- Fine-tuning multimodal and OCR-based VLMs (with PEFT -LoRA, QLoRA, Unsloth) to extract text in a structured manner from historical handwritten and printed industrial records; conducting comparative model analysis.

AI Application Research Project, Computer Vision Institute for Factory Automation and Production Systems (FAPS Lab), Germany **June 2024 - May 2025**

- Developed a multi-modal multi-view multi-task deep learning pipeline for classification of quality monitoring in industrial manufacturing. Github: https://github.com/uddipan77/AI-FAPS-Multi-Modal-View-Task-Pipeline
- A Comparative Assessment of Self- and Semi-Supervised Learning as well as Combined Approaches for Deep Learning Based Image Classification in Industrial Visual Inspection of Electrical Motor Manufacturing. Github: https://github.com/uddipan77/ai-faps-self-semi-combined-dl-pipeline-industrial-inspection

Local LLM-based RAG System for Your Personal Documents

- Built a privacy-preserving local RAG system that indexes my German bureaucracy letters and answers questions using a locally hosted LLM (Ollama) with hybrid search for better recall. I also made a lightweight app that runs entirely on my machine so I can interactively query these documents for my personal use.

  Tools and Technologies used : Ollama, OpenSearch, Hybrid Search, Docker, Streamlit, OCR
  GitHub - https://github.com/uddipan77/local_rag_talk_with_your_docs

Comparing Embedding Techniques for Sentiment-Based Accident Severity Prediction

- Built a sentiment analysis pipeline on news reports to predict accident severity, comparing multiple text embedding techniques—GloVe, Word2Vec, FastText, and Sentence Transformers—paired with a Random Forest classifier.

  GitHub - https://github.com/uddipan77/stock-sentiment-embeddings-comparison

HACKATHON

AINatives Hackathon 2024, Iba GmbH & BEST Erlangen, Germany **June 2024**

Objective:

- Anomaly Detection in time series sensor data using Autoencoders and Isolation Forest
- Development of end to end AI enabled web application using Streamlit

LANGUAGE

English - Fluent **German –** Intermediate **Bengali -** Native **Hindi -** Fluent