ARIJIT SAMAL

+32 493139020 | arijit.samal@student-cs.fr | linkedin.com/in/arijit-samal1 | github.com/1-ARIjitS | devpost.com/arijits19 | 1-arijits.github.io/portfolio

SUMMARY

Results-driven data scientist with big data, machine learning, deep learning, and GEN-AI expertise. Proficient in Python, Big data technologies, PyTorch, and LLMs with a strong background in developing data-driven solutions across various domains, including healthcare, e-commerce, social media, and IoT. Proven ability to design and implement data analysis, NLP, KG, DL, and CV models. Demonstrated success in hackathons, achieved prestigious scholarships, conducted impactful research, and contributed to open-source projects, showcasing strong problem-solving, collaborative, and analytical skills with a commitment to innovate and impact using data science.

EDUCATION

Erasmus Mundus Masters in Big Data Management and Analytics (BDMA) *2023 – Present*

- *Semester 1:* Université libre de Bruxelles (ULB), Brussels, Belgium – MS in Computer Science
- *Semester 2:* Universitat Politècnica de Catalunya (UPC), Barcelona, Spain – MS in BDMA
- *Semester 3:* CentraleSupélec (CS), Université Paris-Saclay, Paris, France – M2 in BDMA

Indian Institute of Science Education and Research (IISER), Bhopal, India *2019 – 2023*

- Bachelors in Electrical Engineering and Computer Science (EECS)
- Minor in Data Science and Engineering (DSE)
- G.P.A.: 9.57 / 10.00

EXPERIENCE

MobilityDB *July – September 2024*
Open Source Developer *Brussels, Belgium*

- Improved JMEOS, Java binding for the MEOS spatiotemporal library *Project Link*
- **Technologies:** C, Java, FFI, CI/CD, GitHub Actions, Python
- Contributed 30K+ lines of code to JMEOS and MobilityDB repositories
- Boosted testing coverage by 70% using JUnit for MEOS data types
- Automated documentation deployment using GitHub Pages, streamlining API visibility for 500+ users
- Built CI/CD pipelines with GitHub Actions, cutting build and integration times by 30%

Health Technologies Lab (HTL), IBME, University of New Brunswick (UNB) *May – August 2023*
Research and Development Intern *Fredericton, Canada*

- Worked on Translating Foot Pressure Maps to 3D Human Poses *Project Link*
- **Technologies:** Pytorch, Python, Mediapipe, TensorFlow, Keras, OpenCV, MATLAB
- Captured foot pressure maps using 100Hz tiles; mapped to 3D poses with 33 keypoints
- Used video from 8 cameras as supervision; developed Encoder-Decoder, CRNN, and CNN+LSTM models
- Evaluated models using MPJPE and MSE, enabling non-invasive person identification with 95% accuracy

SKILLS

- **Languages:** C/C++, Python, Java
- **Programming:** OOPs, Data Structures and Algorithms (DSA)
- **Data Science:** NumPy, Pandas, Seaborn, Matplotlib, Scikit-learn, PyTorch, TensorFlow, Machine Learning, Deep Learning, Data analytics, Data visualization, OpenCV, Langchain, Streamlit, LLM
- **Data Engineering:** SQL, PostgreSQL, Apache Airflow, PySpark, Docker, Shell, GCS, Minio, Big data
- **Graph databases:** SPARQL, Cypher, Neo4J, GraphDB, OrientDB

PROJECTS AND PUBLICATIONS

MediReels *October 2024*
LLM, GEN-AI *Project Link*

- **Technologies:** Python, Mistral-large, GCP, HuggingFace, Streamlit, FastAPI, Edge-tts, Langchain, Moviepy, FLUX.1-dev, Tavily, Asyncio, Pydub
- Mistral X Alan Hackathon project; creating a healthcare platform to simplify trending medical info into engaging short videos (Reels) and long podcasts
- Reduced content creation times from multiple days of research, creation, and editing to under 10 minutes

SpicyBytes *February – July 2024*
BIG DATA, GEN-AI, VLM, LLM, DL, ML *Project Link*

- **Technologies:** Python, Scikit-learn, Selenium, pySpark, MLflow, Streamlit, BigQuery, Minio, GCS, Airflow, Neo4J, GraphDB, Looker Studio, Llama, Langchain, Gemini
- Scraped 1M+ product listings from 15K+ stores across 60+ postal codes of Barcelona
- Platform aimed at reducing food waste for 200K+ students in Barcelona; pitched as a startup at UPC
- Integrated multilingual OCR using Gemini-1.5-pro to automate product inventory from scanned user bills
- Forecasted sales trends using Facebook Prophet and implemented dynamic pricing based on perishability
- Integrated a Llama-based food recommendation engine and BERT-based sentiment analysis for user reviews

Klìnic *May 2024*
LLM, KG, GEN-AI, DL *Project Link*

- **Technologies:** Python, LLM, GPT-4, LangChain, InterSystems IRIS Vector Search, RAG, Streamlit
- 1st place at MLH HackUPC 2024; Platform to assist clinicians and researchers in navigating the landscape of previous clinical trials
- Created a knowledge graph with 500K+ entries from NIH MedGen and clinical trials datasets.
- Enhanced accuracy of queries by using KG embeddings in RAG, leading to an observable reduction in hallucinations while summarizing trials and extracting statistical insights using GPT-4

ThermalVision *December 2022 – August 2023*
IoT, CV, DL, ML *Paper Link*

- **Technologies:** Python, OpenCV, TensorFlow, Keras, PyTorch, Scikit-learn, YOLO, IoT
- Co-authored paper: "ThermalVision: Pioneering Non-Invasive Temperature Tracking in Congested Spaces" as part of my Bachelor's thesis
- Developed real-time temperature tracking in crowded environments using edge devices
- Achieved 85% thermal face detection accuracy and R2 score of 0.96 in real-time temperature estimation

COURSEWORKS

- **Theoretical Computer Science:** Theory of Computation, Computer Organization, Computer Networks
- **Programming:** Introduction to Programming (C/C++), Advanced Programming in Python, DSA
- **Data Science:** Data Science and Machine Learning, Computer Vision, Deep Learning, Data Mining
- **Data Engineering:** DBMS, SQL, Data Warehouse, Advanced Databases, Database System Architecture, Big data management, Semantic data management
- **Mathematics:** Linear Algebra, Discrete Mathematics, Multivariable Calculus

ACHIEVEMENTS

- Selected as 1 of 24 global recipients of the Erasmus Mundus Scholarship for the BDMA Master's program.
- Awarded Overall 1st Place at MLH HackUPC 2024, Europe's largest hackathon.
- Winner of the InterSystems Challenge for Best Use of GenAI and MLH Best Use of MATLAB at HackUPC 2024.
- Received the Mitacs GRI Fellowship (2023); completed research internship at the UNB, Canada.
- Selected as 1 of 50 participants for the Zama X Hugging Face Privacy-Preserving AI Hackathon.
- Achieved Global Rank 108 in the HackerEarth Get a Room ML Hackathon.
- Secured 1st Place in the ArmaCode Alpha 1 Coding Contest at IISER, Bhopal.