

Sayantana Datta

571-438-4947 | sayantandatta275@gmail.com | www.linkedin.com/in/sayantana-datta | github.com/sayantana6720 |

EDUCATION

Northeastern University April 2025
Khoury College of Computer Sciences Arlington, VA
MS in Computer Science GPA (4.0/4.0)

Birla Institute of Technology and Science Pilani July 2022
B.E. in Computer Science CGPA (8.74/10) Dubai, UAE

TECHNICAL SKILLS

Programming languages: Java, Python, Javascript, SQL, R
Python Libraries & Frameworks: PyTorch, TensorFlow, Scikit-Learn, Keras, NumPy, Pandas, Matplotlib, Seaborn, NLTK, FAISS, NLPaug, Hugging Face Transformers
Databases: MySQL, Pinecone
Tools & Platforms: Git, Tableau, Microsoft Excel, Power BI, SPSS, Streamlit, LangChain, OpenAI
AWS: Lambda, Textract, S3, Bedrock, API Gateway. Sagemaker
Certificates: AWS Certified Cloud Practitioner (Aug 2024)

EXPERIENCE

Data Science Intern December 2022 – March 2023
Corizo Bangalore, India

- Cleaned and structured **50K rows** of user data, reducing inconsistencies by **30%** and streamlining analysis
- Engineered features from course feedback and logs to support retention models
- Optimized logistic regression and random forest models to predict user drop-off, improving precision by **15%**

Data Analyst Intern January 2022 – July 2022
Gulf Marketing Group Dubai, UAE

- Conducted attrition analysis on **1,600+ employees** using Python for data manipulation and cleaning
- Performed EDA on employee data across **7 countries and 10 business units** to identify attrition patterns
- Created detailed Excel reports with pivot tables, charts, and conditional formatting to highlight trends
- Assisted in developing retention strategies that led to a **7% reduction** in overall attrition rates

IT Intern May 2020 – August 2020
Power Grid Corporation Delhi, India

- Analyzed outage data using Python (pandas, NumPy) to identify trends in downtime and energy losses
- Created Tableau dashboards to visualize transmission efficiency and load distribution metrics
- Streamlined reporting processes by automating data cleaning tasks, reducing manual effort

PROJECTS

Emotion Detection and Text Transformation System | *XLM-roBERTa, Modal Labs, LoRA* April 2025

- Fine-tuned XLM-roBERTa on XED Dataset with **73,000 examples** and **8 emotion categories** across **10 languages**, winning **second place at Khoury research showcase '24**
- Achieved **0.31 validation loss** by implementing LoRA fine-tuning
- Built a scalable Modal Labs cloud training pipeline, overcoming dependency challenges in distributed environments

SafeNest - Personalize Real Estate Search | *RAG, Vector Search, Pinecone, GPT-4* October 2024

- Developed a property search app using geolocation, crime data visualization, and Haversine distance, **winning first place at aiXplain AI hackathon**
- Enhanced property descriptions with AiXplain models and stored metadata in Pinecone DB for efficient querying
- Built an automated system integrating OpenAI, Pinecone, and with Streamlit for seamless property searches

YouTube InsightBot: AI-Powered Video Q&A | *RAG, pinecone, FAISS, GPT-3.5-Turbo* June 2024

- Built an AI-powered YouTube search system using Retrieval-Augmented Generation (RAG)
- Stored video transcripts in Pinecone DB and used cosine similarity for efficient retrieval
- Integrated LangChain and OpenAI's GPT-3.5-turbo for accurate video-based Q&A