

Debangshu Sadhukhan

7439477347 debangshu.data@gmail.com LinkedIn GitHub

Summary

M.Sc. Data Science student specializing in NLP, Deep Learning, and ML. Skilled in Python, Transformers, and RAG. Passionate about building scalable, production-ready LLM-driven systems.

Education

M.Sc. Data Science & AI (SGPA: 8.80) Ramakrishna Mission Residential College Narendrapur(Affiliated to University of Calcutta)	2024 – 2026
B.Sc. Statistics (CGPA: 9.10) Bidhannagar College	2021 – 2024

Technical Skills

Programming: Python, R, SQL, PL/SQL, Hadoop

Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Power BI, Excel

ML/DL: Regression, Classification, CNN, LSTM, Autoencoders, GANs, Transformers, LLMs, GenAI, RAG, Agents

Tools/Deployment: Streamlit, FastAPI, FAISS, Astra DB, AWS EC2, Docker, Git

Experience

Research Intern – Machine Intelligence Unit, ISI Kolkata

2025

- Analyzed 10K+ colorectal cancer records using survival analysis methods for recurrence forecasting.
- Engineered an ML-driven chart system for cancer image/PDF detection, improving processing efficiency by 30%.
- Streamlined preprocessing workflows, reducing data preparation time by 40%.

FastAPI File Management Intern, IDEAS (ISI Kolkata)

2025

- Developed FastAPI service for CSV/JSON upload, merging, and persistence using PostgreSQL.
- Implemented UUID-based cache + error handling, reducing API latency by 35%.

Projects

Smart Yoga Analyzer — Python, Mediapipe, Random Forest, MLP, CNN

[GitHub]

- Developed real-time yoga posture correction system using Mediapipe + MLP classifier for joint-based pose recognition.
- Achieved 98% pose classification accuracy; angle scoring improved corrective feedback precision.
- Business Impact: Improved posture correctness and reduced instructor workload through automated visual guidance.

Healthcare Chatbot (Groq LLaMA RAG) — Python, RAG, LangChain, AstraDB, Amazon EC2

[GitHub]

- Built medical Q&A chatbot using Groq LLaMA RAG pipeline with AstraDB and EC2 deployment.
- Achieved 25% improvement in retrieval accuracy through optimized chunking, embeddings, and prompt engineering.
- Business Impact: Automated ~900 queries/day, significantly reducing manual support workload.

Vision-to-Language Image Captioning System — Multimodal AI (Computer vision + NLP)

[GitHub]

- Built CNN-LSTM encoder-decoder architecture using VGG16 transfer learning for image-to-text generation.
- Optimized caption quality using sequence modeling, achieving strong BLEU scores with coherent outputs.
- Business Impact: Automated image tagging and enhanced accessibility for scalable content platforms.

Stock Price Movement Prediction using Financial News Sentiment -NLP, BERT, Naive Bayes

[GitHub]

- Designed end-to-end NLP pipeline using BERT/FinBERT for sentiment extraction from financial news.
- Mapped sentiment probabilities to market direction signals with emphasis on downside risk detection.
- Outcome: Linked unstructured news data to structured market intelligence.

SMOTE-Enhanced Fraud Detection & Recall Optimization — Python, Scikit-learn, Streamlit

[GitHub]

- Built fraud detection pipeline using Decision Trees with SMOTE on 284K+ credit card transactions.
- Improved fraud recall to 92% while maintaining 98% accuracy.
- Business Impact: Reduced financial risk through early fraud identification and proactive mitigation.

Certifications

- IBM SkillsBuild – Agentic AI Architect (2025)
- Deloitte Data Analytics Simulation – Forage (2025)
- SQL Bootcamp – Udemy (2025)

Achievements

- 2nd Place – Analysis 2025 (Data Visualization), St. Xavier's College
- Football Champion – Inter-departmental Tournament