

SATYAM DUBEY

+1 (930) 904-6355 | satyamdubey5302@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Education

Indiana University Bloomington

May 2027 (Expected)

Master of Science (MS) – Data Science (Specialization: Big Data Systems)

Courses: Statistics for ML, Cloud Computing Systems, Applied Machine Learning

University of Mumbai, India

Aug 2020 - May 2024

Bachelor of Engineering (BE) – Artificial Intelligence and Data Science

CGPA: 3.7/4

Courses: Deep Learning, Reinforcement Learning, Recommendation Systems, Big Data Technologies

Skills Summary

- **Languages:** Python, C/C++, SQL, Matlab, R, LaTeX
- **Frameworks:** PyTorch, Keras, Scikit-Learn, Transformers, NLTK, Gym, Langchain, Gradio, Flask, PySpark, MLflow
- **MLOps:** Docker, AWS (EC2, SageMaker, Lambda, S3), Firebase, Pinecone, MongoDB, Git, Kafka, AutoGen

Professional Experience

Jio Institute

Ulwe, India

Data Scientist

July 2024 – June 2025

- Created a self-supervised learning pipeline using spectral features and attention for USG representation learning.
- Achieved 5–7% improvement in diagnostic accuracy across multi-organ datasets by enhancing feature generalization.
- Collaborated with clinicians at 3 government hospitals to curate and analyze over 5,000 patient cases, identifying dietary patterns related to fatty liver disease.

Ernst & Young (EY)

Mumbai, India

Summer Analyst

June 2023 – Aug 2023

- Developed the backend for a speech-to-text pipeline with speaker diarization on GCP, enabling agent scoring through voice-based profiling with 85%+ segmentation accuracy.
- Crafted a tone analysis module using speech embeddings and linguistic markers to detect misselling in 1,200+ calls.
- Enhanced RCM (Risk Control Matrix) using a fine-tuned RoBERTa model, reducing compliance review time by 30% via automated risk signal extraction.

Indian Statistical Institute

Kolkata, India

Research Intern

May 2022 – July 2022

- Implemented color normalization for H&E stained images and evaluated RGB, HSI, and Lab color models.
- Found HSI channel to yield superior accuracy for breast cancer cell analysis through comparative model assessment.s

Projects Experience

DocSpot [\[GitHub\]](#) | ElasticSearch, FAISS, Langchain, MongoDB, React, Flask

May 2024

- Built multilingual academic research assistant combining Gemini-powered reasoning with RAG architecture for document insights, real-time chat, and translation.
- Reduced average search latency by 43% by combining ElasticSearch keyword matching with FAISS semantic search.
- Implemented T5-small summarization pipeline transforming complex papers into digestible insights for faster research.

SecureGans [\[GitHub\]](#) | GANs, Flask, React, U-Net

Apr 2024

- Designed a full-stack web system to restore occluded or masked facial features using a GAN-based inpainting model.
- Achieved PSNR of 22.25 and SSIM of 0.874 with dual-path U-Net ensemble, leading to publication in JES.

Connect-4 AI [\[GitHub\]](#) | Python, DQN, RL

Jan 2024

- Explored exploration-exploitation trade-offs in Deep Q-Learning by implementing 2 strategies: epsilon-greedy, UCB.
- Built evaluation framework comparing 5+ DQN variants against minimax baselines, achieving 78% win rate.

Publications

- Satyam Dubey, Tanushree Meena, Dwarikanath Mahapatra, Behzad Bozorgtabar, Sudipta Roy, “Multi-Band Spectral Subspace Learning for Domain-Invariant Ultrasound Representation”, (under review 2025).
- Satyam Dubey, Jagannath Nirmal, “Beyond Texture: Unveiling Spiny Crown-of-Thorns Starfish with Multiresolution Analysis”, *Journal of Intelligent Marine Technology and Systems*, Springer, vol. 2, no. 1, p. 16, 2024. [\[Link\]](#)
- Anjali Yeole, Prathmesh Pawar, Satyam Dubey, Yash Sarang, Arunim Chakraborty, “Image Inpainting for Missing Facial Data Recovery in Security Settings”, *Journal of Electrical Systems*, vol. 20, no. 3, pp. 3165–3171, 2024. [\[Link\]](#)

Achievements

- Winner, Hackathon 'WELDRIGHT' conducted by IIT-B & Godrej Aerospace, 2023. We built a CI/CD-based welding defect detection system using visual analysis and sensor inputs for real-time monitoring.
- Winner in *HACK-AI-THON 2.0* in the Elastic Search track, conducted by AI CoLegion and Elastics, 2024.