

NITIN KUMAR JHA

Data Scientist

Phone No: +91-7691884063 **Email Id:** nkjha877@gmail.com **Address:** Chennai, IN
LinkedIn: <https://www.linkedin.com/in/nitin-jha-165180191/>

Results-driven Data Scientist with a strong foundation in mathematics and hands-on experience in machine learning, deep learning, computer vision, and GenAI. Seeking a challenging role in a data-driven organization to apply analytical expertise, programming proficiency, and problem-solving skills to extract insights from complex datasets.

PROFESSIONAL SUMMARY

Innovative and research-driven Data Scientist with 5+ years of hands-on experience in solving real-world problems through Deep Learning, Computer Vision, and Generative AI. Proficient in building and deploying scalable AI solutions—ranging from OCR-based behavioral analytics and object detection pipelines to LLM-powered applications using OpenAI, Hugging Face Transformers, and LangChain. Adept at combining structured and unstructured data with advanced neural architectures to drive impact in domains like education, and NLP. Passionate about applied AI research, with strong foundations in model building (ML, DL, LLM), and end-to-end pipeline development using tools like Sklearn, PyTorch, Huggingface.

SKILLS

Technical Skills: Python, PyTorch, scikit-learn, SQL, LangChain, LangGraph, openAI API, PowerBI, Tableau, Plotly, matplotlib, data visualization, RAG Pipelines, Pandas, NumPy, Image preprocessing, Object Detection, Image Classification, OCR & Text Extraction, OpenCV, Hugging Face, Fine-Tuning models(LoRA, QLoRA)

Core Competencies: Mathematical & Statistical Modeling, Research-driven Model Development (ML, DL), Generative AI and Large Language Models (LLMs), Multi-modal Deep Learning (CV + NLP), OCR & Screen Analytics for Behavior Inference, Project Mentorship, AI Curriculum Development & Instructional Design, Cross-functional Collaboration

Soft Skills: Problem-solving, Communication, Mentorship, Collaboration, Adaptability

WORK EXPERIENCE

Project Associate

IIT Madras (Chennai (In-person))

- Built a computer vision pipeline with OCR and YOLOv8 to detect browser tab-switching in 100K+ screen captures during proctored exams.
- Modeled student behavioral patterns from quiz audit logs using transition matrices, clustering, and time-based features.
- Developed a domain-agnostic AI analytics platform, facilitating scalable AI solutions for diverse MSME manufacturing segments.
- Implemented a Deep Learning curriculum emphasizing mathematical foundations and real-world applications using PyTorch.
- Created structured hands-on labs for prompt tuning, instruction following, and JSON-based output parsing using OpenAI + LangChain.
- Revamped Machine Learning Techniques course to prioritize hands-on implementation of regression, classification, and clustering algorithms.
- Redesigned Statistics course, enhancing conceptual clarity on probability, distributions, hypothesis testing, and estimation.
- Integrated scikit-learn-based exercises in the Machine Learning Practice course, promoting applied learning and practical exposure.
- Revamped advanced courses in LLMs, GenAI, NLP, and Computer Vision by integrating mathematical foundations like attention mechanisms, loss functions, and representation learning.
- Enabled hands-on implementation of transformer architectures, generative models (autoregressive, diffusion), and RAG systems using PyTorch and Hugging Face.
- Facilitated advanced-level sessions, providing technical guidance and support to a diverse student base.
- Collaborated with academic and technical teams to align content with evolving industry trends.

Achievements:

- Enhanced learning outcomes for 20,000+ students through redesigned courses.
- Led data analysis and curriculum design teams, driving the development of projects and technical course content.
- Boosted student engagement by 65% through the integration of hands-on scikit-learn labs.
- Improved implementation success by 80% through PyTorch-based deep neural network workshops for 300+ participants.
- Enhanced comprehension and applied skills by delivering targeted training on machine learning fundamentals with NumPy to 1000+ students.

Aug 2020 - Present

EDUCATION

Master of Science (M.Sc.) in Mathematics

Indian Institute of Technology, Madras
(Aug 2018 - May 2020)

Bachelor of Science (B.Sc.) in Mathematics

Central University of Rajasthan, Ajmer
(Aug 2015 - May 2018)

ACHIEVEMENTS

- AIR 65 - CUCET 2015: Secured a top rank in the Central Universities Common Entrance Test, reflecting academic excellence and competitive merit.
- Institute Merit Scholarship - IIT Madras: Awarded for outstanding academic performance and commitment to continuous learning.
- AIR 126 - JAM 2018: Ranked among the top nationwide in the Joint Admission Test for M.Sc., showcasing strong mathematical aptitude.
- INSPIRE Award - DST, 2010: Recognized by the Department of Science and Technology for innovation and scientific talent at a national level.

Projects

Tab Switching Detection and Behavior Modeling During Online Exams

- Built a computer vision pipeline to detect browser tab-switching behavior from time-stamped screen captures collected from over 20,000 students across 10+ Online Proctored Programming Exams (OPPE), using OCR and image preprocessing to analyze tab navigation patterns.
- Modeled student navigation timelines by reconstructing tab sequences, computing switch frequency, time spent per tab, and generating behavioral tags (e.g., focused and suspicious).

Deep Learning-Based Detection and Classification of Mosquito Species

- Developed an end-to-end deep learning pipeline to detect and classify mosquito species from high-resolution images using YOLOv8 for object detection and ResNet50 for species-level classification.
- Applied data augmentation, transfer learning, and cross-validation, achieving high precision and recall on an imbalanced dataset of mosquito images.
- Evaluated model performance using mean Average Precision (mAP), tuning IoU thresholds and detection confidence to maximize precision-recall trade-off for mosquito detection accuracy.

Multilingual Sentiment Analysis using LLaMA 3.1-8B

- Engineered parameter-efficient fine-tuning of LLAMA 3.1-8B-Instruct for sentiment classification, encompassing 13 Indian languages.
- Adapted the model's tokenizer and training data to facilitate multilingual capabilities, extending beyond officially supported languages.
- Implemented LoRA-based parameter-efficient methods to boost performance in low-resource language scenarios.

Multilingual Document Assistant

- Designed and implemented a multilingual Retrieval-Augmented Generation (RAG) pipeline using LangChain for querying statutes, case laws, and regulations in multiple languages (e.g., English, Hindi, Tamil).
- Built a hybrid retrieval system (dense + BM25) with a reranking stage to improve precision in legal text retrieval.
- Integrated multilingual embeddings (e.g., LaBSE, E5-Multilingual) to enable cross-language querying without translation loss.
- Developed citation-grounded Q&A chains in LangChain to ensure all answers referenced exact legal sections and maintained compliance disclaimers.

Audit Log Sequence Mining Project - IIT Madras

- Directed advanced sequence pattern mining on audit logs from online examination platforms to detect behavioral anomalies.
- Analyzed audit logs from 30+ exam days (60,000+ records) using Python, Pandas, and Plotly for enhanced data preprocessing and visualization.
- Employed sequence alignment and clustering algorithms (K-Means, DBSCAN) to discover recurring patterns and irregularities in student interaction sequences.
- Deployed ML/DL models to predict student performance in different courses.