

SUMMARY

Data Science graduate student with hands-on experience in AI-driven solutions and a commitment to continuous learning. Proven track record of deploying innovative technologies in real-world applications and contributing to academic research.

EDUCATION

University of California San Diego <i>Master of Science, Data Science</i> Teaching Assistant for DSC-261: Data Ethics, DSC-291: Statistical Models	September 2022 – December 2024 GPA (3.98/4.0)
Delhi Technological University, New Delhi <i>Bachelor of Technology, Mathematics and Computing Engineering</i>	August 2018 – June 2022 CGPA (8.73/10)

WORK EXPERIENCE

Analytical Scientist Intern <i>FICO, San Diego, California</i>	June 2023 – December 2023
<ul style="list-style-type: none">Devised and implemented adaptive time-series algorithm to monitor the latent features of a State-of-the-Art fraud detection neural network and trigger real-time alerts for significant shifts in distributions. Validated the algorithm for 15 major clients.Developed an extensive pipeline to compute and visualize the distributions of terabyte-scale transaction datasets using PySpark.Conducted data simulations to emulate drastic shifts in customer behavior patterns and cluster new fraud schemes using t-SNE.	
Research Engineer <i>Collablens, Haryana, India Funded by MIT Media Lab</i>	January 2022 – September 2022
<ul style="list-style-type: none">Developed and deployed an AI station for quality assurance drop testing of flour packets. Integrated dynamic cloud-based modules for Spillage Detection, Failure Rates, and other real-time insights from live video footage.Helped secure a contract to deploy the system in 50 factories. Helped raise over \$200,000 in investment offers.Prototyped a versatile Computer Vision System to detect various types of defects in laser-engraved wooden boards on a moving assembly line, flagging defective units and storing results on the cloud. Achieved mean inference time of 2.5 seconds per board.	
Machine Learning Intern <i>Hypertechpreneurs, Haryana, India Funded by Oriental Insurance</i>	May 2021 – December 2021
<ul style="list-style-type: none">Developed and productionized Vehicle Damage Detection Model utilizing Mask R-CNN for Instance Segmentation to automate vehicle inspections. Extrapolated it to a Severity and Cost Estimation pipeline. Helped raise over \$50,000 in funding.Developed systems for OCR and Object Detection in dynamic environments while maintaining a minimum accuracy of 90%.	

RESEARCH EXPERIENCE AND PUBLICATIONS

Lead Research Fellow under Prof. H.C. Taneja, Delhi Technological University	September 2021 – May 2022
<ul style="list-style-type: none">Outperformed the Black-Scholes Model for option pricing using LSTM, MLP, XGBoost and SVM leveraging real market data.Sood, S., Jain, T., Batra, N., Taneja, H.C. (2023). Black-Scholes Option Pricing Using Machine Learning. In: Proceedings of International Conference on Data Science and Applications. Lecture Notes in Networks and Systems, vol 551.	
Research Assistant under Prof. Anurag Goel, Delhi Technological University	February 2021 – August 2021
<ul style="list-style-type: none">Integrated State-of-the-Art CNN based Object Detection Networks (CenterNet, Faster R-CNN) with self-devised algorithms for selective lossy image compression techniques to enhance the storage and processing efficiency in autonomous systems.S. Sood and Y. Ahuja, "Selective Lossy Image Compression for Autonomous Systems," 2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2021, pp. 1-5.	

PROJECTS

Rubik's Cube 3D visualizer and solver using MCTS	August 2024 – Present
<ul style="list-style-type: none">Developed a 3D visualizer for the Rubik's Cube, with ongoing work on integrating Monte Carlo Tree Search (MCTS) and reinforcement learning to optimize the solving process without human input.	
MediLoRA: LLM for medical Q&A with QLoRA	October 2023 – January 2024
<ul style="list-style-type: none">Fine-tuned OpenHermes-2.5-Mistral-7B with Q-LoRA on 300M medical text tokens. Improved PubMedQA and MedQA accuracy by over 20% and matched State-of-The-Art 70B open models on MMLU-Medical with 0.05% of the data size.	

TECHNICAL SKILLS

Programming	: Python, R, MATLAB, C++, JavaScript, HTML, CSS
Technologies	: Pandas, PyTorch, Keras, TensorFlow, OpenCV, AWS, PostgreSQL, Bash, PySpark, Docker, Airtable
Skills	: Data Science, Data Analytics, Deep Learning, Computer Vision, Financial Modeling, Fraud Detection