

Suraj Thapa

✉ hi.surajthapa@gmail.com | 📞 (203) 627-6793 | 🔗 linkedin.com/in/mgrsuraz | 🐙 github.com/mgrsuraz

EDUCATION

University of New Haven

MS student of Science in Data Science (Tagliatela College of Engineering)

Aug. 2024 - Exp. May. 2026

West Haven, CT

GPA 4.0 / 4.0

- **Concentration:** Machine Learning, Computer Vision
- **Courses:** Mathematics, Data Engineering, Deep Learning, Natural Language Processing

Nepal College of Information Technology

Bachelor of Engineering in Information Technology

Sep. 2017 - Aug. 2022

Kathmandu, Nepal

GPA 3.89 / 4.0

- **Courses:** Data Structure and Algorithms, Engineering Mathematics, Applied Operating Systems, Computer Graphics, Signals Systems and Processing

PROFESSIONAL EXPERIENCE

University of New Haven

Research Assistant (Python, Tmux, LaTeX)

Aug. 2024 - Present

West Haven, CT

- Developed a novel density estimation algorithm for robust model fitting in computer vision, reducing iterations by $\sim 60\%$ vs. RANSAC variants while boosting inlier count by $\sim 6\%$ on benchmark datasets.
- Implemented the algorithm in Python (PyTorch/TensorFlow); ran experiments validating gains in outlier rejection and efficiency.
- Collaborated on ML/CV research projects; contributed to the development of papers on methodologies and results.

Cliff Byte Pvt. Ltd.

Data Analyst (Git, Vim, R)

Aug. 2022 - Dec. 2023

Kathmandu, Nepal

- Processed and cleaned 800K+ row election datasets using SQL, R, and Python (Pandas, NumPy) to identify key trends and patterns.
- Uncovered 15–20% cost-saving opportunities through statistical trend analysis of government election data, enabling data-driven decisions for public-sector stakeholders.
- Built interactive Power BI dashboards visualizing election metrics and insights, accessed weekly by 10+ stakeholders.

SKILLS

- **Languages :** C++, C, Python, Java, R, SQL, HTML, CSS, JavaScript, Matlab
- **Framework / Tools :** Git, Flask, Postman, Linux, Docker, Vim, Tmux, AWS, LaTeX, PowerBI
- **Libraries :** Scikit-Learn, PyTorch, TensorFlow, OpenCV, Numpy, Pandas, Matplotlib

PROJECTS

Instance Segmentation using Mask R-CNN (PyTorch, Mask R-CNN)

Aug. 2025 - Dec. 2025

- Led a team of 3 in end-to-end fine-tuning of Mask R-CNN on a custom dataset of 220 annotated images (pen, plug, and shoe), using COCO-pretrained weights and ResNet-50-FPN backbone.
- Performed hyperparameter tuning and trained for 20 epochs, optimizing multi-task losses (classification, bounding box regression, mask) to achieve robust convergence on limited data.
- Delivered strong segmentation performance: mean IoU ≈ 0.79 (train) / 0.80 (val) for precise pixel-level masks; mAP@0.50 of 0.093 (train) / 0.119 (val) for effective instance detection despite small size of the dataset.

FinRAG: Financial Question Answering using Retrieval-Augmented Generation (Python, NLP, RAG, LLMs)

Aug. 2025 - Dec. 2025

- Designed and implemented a domain-specific RAG pipeline over 290,000 documents (Reddit threads from 6 finance subreddits, 1-year OHLCV for 10 stocks, 16 books/papers) for accurate, grounded financial QA.
- Built retrieval layer with BAAI/bge-large-en-v1.5 embeddings and FAISS index ($>280,000$ chunks); integrated Mistral-7B, Phi-3-Mini, and Llama-3.1-8B for generation.
- Evaluated on 24 investment/market queries: achieved 0.48 semantic similarity across models; Mistral-7B showed best latency (1/3 of Phi-3, 1/2 of Llama-3.1).