

Riya Raut

✉ rraut@binghamton.edu | 📞 6072968176 | 🌐 /in/riyaraut

Education

Binghamton University, State University of New York	MS in Computer Science	Aug 2023 – May 2025
MIT ADT University, School of Engineering	B-Tech in Computer Science	Aug 2019 – May 2023

Technologies

Languages: Python, C, C++, Java, R

AI Tools: Generative AI, Prompt Engineering, Langchain, RAG, Knowledge Graphs, OpenCV, HuggingFace, Sci-Kit Learn, NumPy, Pandas, PyTorch, TensorFlow, Plotly, Matplotlib, Seaborn, ggplot, Tableau, NLP

Databases: MySQL, PL/SQL, MongoDB, Neo4j, Vector Databases, Pinecone, PostgreSQL

Cloud & DevOps: AWS Cloud, Docker, Kubernetes, CI/CD, Git

Web Technologies: HTML, CSS, JavaScript, Django

Work Experience

Machine Learning Intern, Julius – New York City, New York Jan 2025 – Present

- Implemented **Semantic Similarity Search and Clustering methods** to improve tag categorization, eliminate redundancies, and enhance data consistency. This led to **higher-quality training data**, resulting in more accurate and **contextually relevant LLM outputs** and better model performance.
- Conducted **statistical frequency analysis** to identify dominant patterns in tagging data, ensuring a more structured and noise-free dataset which improved feature extraction for downstream ML tasks.

AI Engineering Intern, Virtusa – Milpitas, California Jun 2024 – Aug 2024

- Published a **White Paper** and led a project on **Advertisement Localization** using **Generative AI** tools like **LangChain**, **OpenAI API**, and **Prompt Engineering** to tailor content for regional demographics, utilizing **Gradio** for the front-end.
- Built a **Recommendation Engine** by leveraging **Neo4j Knowledge Graphs** and utilizing the **Hybrid Filtering** technique for accurate recommendations. Employed **Plotly** to visualize the results.
- Analyzed and pre-processed large datasets using **NumPy**, **Pandas**, and performed detailed **Statistical Analysis**.

Machine Learning Intern, Kanverse.AI – Pune, India Jan 2023 – Jul 2023

- Engineered advanced similarity search algorithms using **Pinecone Vector Database**, improving retrieval accuracy and processing speed for large accounting datasets.
- Analyzed and visualized data from **JSON** files generated through **PDF annotations** for a database of over **10,000** documents. Utilized **Python**, **NumPy**, **Pandas**, and **Plotly** to extract valuable insights.
- Spearheaded a **YOLOv5-based Semantic Segmentation** model for automated insurance claims to assess car damage, ensuring efficiency and fairness in the payout process.

Machine Learning Intern, Indian Council of Agricultural Research – Pune, India Nov 2021 – Jan 2023

- Developed CNN and Mask R-CNN models using TensorFlow for early detection of grape crop diseases, boosting yields by 30-40%
- Led a research initiative focused on applying **AI-driven solutions** to promote **sustainable agriculture** through early disease detection.

Data Analyst Trainee, WNS Global Services – Pune, India Aug 2021 – Oct 2021

- Contributed to in-house **machine learning projects** by optimizing data processes using **Python**, **Pandas**, and **SQL**, resulting in measurable improvements in project outcomes.

Projects

Industry-Specific Layoff Tracker Pipeline | *MongoDB, Faktory, Flask, NLTK, Matplotlib* [Link to Project](#)

- Developed an automated data pipeline that scraped and processed **200,000+** records from **Reddit** and **4chan**, leveraging **MongoDB**, **Faktory workers**, and **Flask APIs** for real-time layoff analysis.
- Applied **sentiment and toxicity analysis** using **NLTK**, achieving **95%** accuracy, and created data visualizations with **Matplotlib** to generate actionable industry insights.

Real-Time Sign Language to Text Translator using Deep Learning | *TensorFlow, LSTM, OpevCv* [Link to Project](#)

- Developed a real-time sign language recognition system using **CNN** and **MediaPipe Holistic**, achieving **97%** accuracy in gesture recognition.

Detection of Tuberculosis using Transfer Learning | *TensorFlow, Transfer Learning Models* [Link to Project](#)

- Assessed **InceptionV3**, **EfficientNetB3**, **DenseNet201**, and **ResNet50** models for tuberculosis detection using chest X-rays, achieving an impressive **99.95%** accuracy on the **TBX11K** dataset, contributing to **2.4M+** diagnoses across the country.

Certifications & Publications

- AWS Certified Solutions Architect – Associate** [Link](#)
- IEEE Publication:** A Comparative Study of Detection of Tuberculosis using Machine Learning and Deep Learning [Link](#)