Rithesh Kumar

(831) 266-7201 | rkumar45@ucsc.edu | rithesh-kumar-91332a156 | github.com/Rithesh17 | rithesh17.github.io

Education

University of California, Santa Cruz | **Santa Cruz, CA** — *September 2023 – June 2025*
Master of Science in Computer Science — ***GPA: 3.9/4.0***

National Institute of Technology Karnataka | **Surathkal, India** — *August 2011 – May 2015*
Bachelor of Technology in Computer Science — ***GPA 9.4/10.0***

Experience

Accretional — San Francisco, CA
Artificial Intelligence Engineering Intern — *June 2024 – August 2024*

- Built **Brilliant**, a GenAI platform for automated API generation using LLMs and RAG pipeline, enabling deployment to AWS Lambda and GCP Functions.
- Engineered advanced prompt workflows and integrated vector databases for semantic search, improving query relevance by 20% and reducing hallucination rates by 30%.
- Collaborated with product and engineering teams to refine RAG pipelines, enhancing API generation accuracy and deployment reliability.

Goldman Sachs — Bengaluru, India
Senior Machine Learning Scientist — *January 2023 – July 2023*

- Led a small team of ML Engineers to develop and deploy real-time machine learning pipelines for credit risk scoring and segmentation, reducing processing latency by 15% and supporting personalized decisions for over 1M users monthly.
- Spearheaded the integration of ML services into the Provenir backend, reducing deployment time by 25% and enabling seamless A/B testing workflows across the personal loans division.
- Drove cross-functional collaboration with product and infra teams to scale credit decision systems, improving deployment reliability and decreasing maintenance overhead by 20%.

Machine Learning Scientist — *August 2020 – December 2022*

- Designed and trained LSTM-based models for time-series analysis of customer repayment behavior, resulting in a 12% improvement in delinquency prediction accuracy.
- Conducted risk modeling and scenario simulations to study COVID-19's financial impact, contributing to a 10% increase in new customer acquisition through targeted lending strategies.
- Developed SVM classifiers to support delinquency forecasting, contributing to reduced defaults and improved credit decision efficiency.

Sprinklr India — Gurgaon, India
Machine Learning Research Intern — *May 2019 – July 2019*

- Improved sentiment analysis accuracy by 10% by optimizing LSTM-based models used for analyzing customer reviews from social media platforms.
- Researched and applied model compression techniques to reduce LSTM model size by 60% while maintaining comparable accuracy, enabling deployment on resource-constrained systems.

Publication

Prostate Cancer Grading Using Multistage Deep Neural Networks, A novel multi-stage deep learning framework for automated Gleason system grading (GSG) of prostate cancer cells. DOI: 978-981-19-5868-7_21

Projects

CT Deniosing and Explainability using OT-CycleGAN — *January 2025 – Present*

- Adapted the OT-CycleGAN architecture to denoise ultra-low-dose CT scans from the IEEE COVID-19 CT dataset.
- Introduced Grad-CAM, attention maps, and PSNR/SSIM overlays to assess denoising transparency and aid clinical interpretability.

Additional Projects: Whispers of the Heart: AI Therapy Assistant (Python), Academic Lecture Video Summariser (Python), GeneWeaver: A Parametric Hardware Generator for DNA Sequence Alignment (Scala)

Skills

Languages: Python, C++, SQL, JavaScript
Machine Learning Libraries: PyTorch, TensorFlow, Scikit-learn, HuggingFace, OpenCV
Generative AI Techniques: Diffusion Models, GANs, Prompt Engineering, RAG
Cloud & Deployment Tools: AWS Lambda, Google Cloud Functions, Docker, Vertex AI, FastAPI, REST APIs
Data Infrastructure: PostgreSQL, MongoDB, DynamoDB, Redis, Kafka, CloudWatch