

# PRATIK DOSHI

+1(831) 266-8773 ♦ Santa Cruz, CA

[prdoshi@ucsc.edu](mailto:prdoshi@ucsc.edu) ♦ [linkedin.com/in/pratik-doshi-b2a493153/](https://www.linkedin.com/in/pratik-doshi-b2a493153/) ♦ [github.com/Pratik-Doshi-99](https://github.com/Pratik-Doshi-99)

## OBJECTIVE

---

Machine Learning focused Graduate Student seeking Machine Learning/AI roles.

## EDUCATION

---

**MS Computer Science**, University of California, Santa Cruz Expected: 2025

Relevant Coursework: Applied ML, Deep Learning, Compilers, Linear Algebra and Artificial Intelligence.

**Post-Graduate Diploma in Data Science**, Mumbai University 2022 - 2023

Relevant Coursework: Machine Learning, NLP, Statistical Methods, and Python.

## SKILLS

---

<b>Machine Learning</b>	LLMs, RAG, Fine-tuning, Classical ML, Deep Learning, Distributed Training and Inference
<b>Technical Skills</b>	PyTorch, Python, Tensorflow, C, Ollama, vLLMs, Kubernetes and Docker
<b>Soft Skills</b>	Technical Writing, Documentation, Communication, Reporting and Presentation skills

## EXPERIENCE

---

**Research Intern** July 2024 - Present

Data Care LLC, Utah, USA

- Using Colossal AI, Slurm and Kubernetes for distributed training on NVIDIA A6000 and A4000.
- Doing research on cutting-edge frameworks and optimization techniques for low-latency LLM inference.

**Associate Software Engineer** June 2021 - Mar 2023

Rupeeseed Technology Ventures, Mumbai, India

- Reduced turnaround latency for a recommendation system from 15 minutes to 2 seconds using LINQ in C#.
- Designed a data processing pipeline in C# and improved its throughput by 50% using pipeline parallelism.
- Developed a statistical model to predict the profitability of customized trading strategies using probability distributions and derivative pricing models.

## PROJECTS

---

**Job Description Summarization using Llama 3.** Built an AI Agentic workflow using Llama 3-70B to fetch and summarize job descriptions associated with a given role. Compiled a dataset of 165 jobs and required skill sets, performed EDA, and published results on LinkedIn. ([Github](#))

**Image Captioning using VLMs.** Achieved 25% accuracy improvement over LSTM baseline on the BLEU metric by using dynamic attention mechanism from the paper "Show, Attend and Tell". In this project, I trained small Vision Language models on image-to-sequence tasks. ([Github](#))

**Analyzing Sparse Autoencoders (SAEs) using Linear Probing.** Trained an SAE on Tiny-Stories-21M (base model) and used Linear Probing to compare inner representations on a synthetic dataset. ([Github](#))

**Trained a small LM.** Trained a small LM. Wrote an Auto-regressive training loop in Python (PyTorch) for training a small Language Model on a multi-GPU Kubernetes Cluster.

**Research on Emissions.** Published a paper on a novel system to estimate emissions-related externalities of power plants by factoring population density in the Gaussian Plume Dispersion Model.

## LEADERSHIP

---

- Served as the President of the Tech Club of ASMSOC, NMIMS. Grew the club from 60 to over 200 members and developed "EventSync", a community-driven mobile app to keep students updated on college activities.