

Sai Paresh Karyekar

🏠 Atlanta, GA 📞 (404) 200-3907 ✉️ saikaryekar@gmail.com 🌐 [saikaryekar](#) 📄 [sai-karyekar](#)

EDUCATION

Georgia Institute of Technology, Atlanta, GA

May 2025

MS in Electrical and Computer Engineering (Machine Learning)

GPA: 3.85/4

University of Mumbai, India

June 2023

B.Tech in Electronics & Telecommunication Engineering

CGPA: 9.62/10

RELEVANT COURSEWORK

Generative and Geometric Deep Learning, Data and Visual Analytics, Regression Analysis, Machine Learning, Digital Image Processing, Technology Entrepreneurship (**Teaching Assistant**)

EXPERIENCE

NVIDIA | *Developer Marketing Intern (Product Marketing)*

June 2024 - Aug 2024

- Developed a Generative AI powered internal tool for end-to-end technical blog generation using the Llama3-70b-instruct model hosted on NVIDIA NIM API Catalog.
- This tool employs Retrieval-Augmented Generation (RAG) with the FastEmbed Embedding model, using nvidia/rerank-qa-mistral-4b for reranking and FAISS for local vector embedding storage.
- Implemented advanced prompt engineering techniques to generate comprehensive blog outlines, detailed content, and corresponding social media posts from provided documents and URLs.

Georgia Institute of Technology, Atlanta | *Student Researcher* 🌐

Aug 2023 - Dec 2023

- Applied statistical techniques (ANOVA, OLS regression) and machine learning models to analyze a large-scale dataset of student performance in a microelectronics course.
- Developed predictive models using scikit-learn and performed feature importance analysis, leading to a 5% improvement in course success rates.

PROJECTS

Enhancing Mathematical Reasoning in Small Language Models (SLMs) 🌐

Sept 2024 - Dec 2024

- Optimized the T5-small transformer model (60M parameters) for math reasoning on the GSM8K dataset using techniques such as Low-Rank Adaptation (LoRA), Chain-of-Thought (CoT) prompting, and full fine-tuning.
- Compared CoT-based methods to a quantized T5-base model, demonstrating competitive performance with lower computational costs.

ApplyCation: Automated Job Application Platform 🌐

Oct 2024

- Built an automated job application system using Selenium and dynamic script generation via Claude API, streamlining the application process by automatically filling out job application forms on various job portals.
- Developed a responsive frontend interface with Streamlit to manage user settings, upload resumes, and track job applications, enhancing the user experience and accessibility.

ML based Portfolio Management with Macro-Financial Indicators 🌐

Jan 2024 - May 2024

- Developed an ML-based portfolio management system integrating macro-financial indicators to optimize portfolio allocation. Utilized various ML models, with DBSCAN achieving a high Silhouette score of 0.737
- Implemented K-means, ARIMA, and LSTM models to predict stock prices, resulting in improved predictions and a higher Sharpe ratio.

SKILLS

- Programming Languages:** Python, C++, SQL, R, MATLAB
- Tools:** Git, Linux, GCP, AWS, Docker
- Frameworks:** Tensorflow, PyTorch, Scikit-learn, OpenCV, LangChain, LlamaIndex, Selenium, Streamlit

EXTRA-CURRICULARS

- As a member of the Data Science club at GT, contributed to the Workout of the Day prediction project by estimating workout durations and comparing performance with CrossFit enthusiasts.
- Social Chair of the Grad Society of Women Engineers (Grad SWE).