SAMARTH K J

Email | LinkedIn | Github

EDUCATION

Indian Institute of Technology, Madras
BTech, Civil Engineering, CGPA: **8.86**   *Jun, 2026*

Kendriya Vidyalaya, Malleshwaram
Grade 12, Percentage: **95.2**   *May, 2022*
Grade 10, Percentage: **96.0**   *May, 2020*

SCHOLASTIC ACHIEVEMENTS

- Secured **99.46% ile in JEE(M)** among 1M+ students
- Achieved **AIR 1473 in KVPY** among 50K+ students
- Secured **AIR 5957 in JEE(A)** among 150K+ students
- Scored **324/390 in BITSAT**, was offered BE in CS at BITS Pilani

RELEVANT COURSEWORK

- Deep Learning
- ML Foundations
- DL in Speech*
- Linear Optimization
- Convex Optimization*
- Probability & Statistics
- Differential Equations
- Series and Matrices

EXPERIENCE

*Ongoing

Fractal AI Research | AI Research Intern   *May 2025 – Aug 2025*
- Finetuned **SigLIP2** using **DeepSpeed ZeRO** with a **novel aesthetics-based dataset** for reward modelling.
- Orchestrated **cross-attn and CFG-based guidance methods** to improve image fidelity of DiTs by $\approx$ **10%**.
- **Curated OCR data** to finetune DiT using **Florence-2** with **GRPO**; increased text-gen accuracy by **24.5%**.

Massachusetts Institute of Technology | Student Researcher   *Jan 2025 – May 2025*
- Built an **agentic system** to aid **memory** with **AI**; submitted to the **flagship conference ACM CHI 2026**.
- Built a **novel data pipeline** on Python FastAPI server **reducing the latency by 80%** using concurrency.
- Evaluated transcripts with **WER and CER**, achieved **4.53% and 1.54%**. Used **LLMs** to flag unsafe outputs.

Dirac Labs | ML Associate   *Feb 2024 – Dec 2024*
- Implemented a **SciML-based NN** in the Julia language for real-time edge applications at <**5ms inference rate**.
- Strategised a data pipeline on **Google Cloud** for the pre-processing, transformation & storage for >**5TB** of data.
- Built an **RF model** to detect soil type over ground using hyperspectral satellite data with an **accuracy of 95.7%**.

PROJECTS

Efficient Tamil ASR with LoRA | *PyTorch, Transformers, PEFT, KenLM*   **Aug 2025**
- **LoRA**-tuned IndicWav2Vec (a **317M** params model) on **201-hrs** of Tamil speech, yielding a **13.45% WER**.
- Cut trainable parameters by **99%+** and model storage by **99.3% (1.2 GB to 8 MB)**, enabling scalable ASR.
- Used a **5-gram** Tamil KenLM with beam search, **boosting accuracy by 2.3%** over a greedy-decoding baseline.

StrInt: Structural Intelligence | *LangGraph, LangChain, Streamlit, ragas, FlagEmbedding*   **Aug 2025 – Present**
- Built an **agentic AI system**, using **LangGraph** & **Gemini API** to design structures acc. to Indian Standards.
- Improved context relevance by **23.4%** via **reranking**, **smart-chunking** & extracting with **MLLM-based OCR**.
- Achieved **0.96 faithfulness**, **0.92 answer relevancy** and **0.79 context relevancy** with **ragas** evaluation.

Earthquake Response Spectrum Prediction | *PyTorch, Pandas, MatPlotLib*   **Jan 2024 – May 2024**
- Built **GenAI models** like **CGAN & VAE** to predict Structural Impact of Earthquakes using 6 EQ params.
- Implemented a **novel-loss** function for **CGAN** in PyTorch. Improved the test-data R^2 value to a SoTA **0.91**.
- Used XGBoost, FFNNs and BNNs to predict ground acceleration values to get R^2 **values of 0.82, 0.84 & 0.87.**

Credit Card Fraud Detection | *Pandas, Scikit-Learn, Optuna*   **Aug 2024**
- **F1** score of **0.89** was achieved with **CatBoost** to detect fraudulent accounts using data from **608K+** accounts.
- Data preprocessing techniques including **SMOTE, ANOVA, Chi-Square test**, etc. were used before modeling.

COMPETITIONS

- Secured **1st place** in Buildathon; used **Flutter** to create **OCR & SQLite** based data digitization application.
- **Headed a 5-member team** & **devised strategy** to bag **1st place** in Mock Auction among **345** participants.
- Devised innovative AI solutions to win **3rd place pan-India** in **L&T** Smart City Challenge among **180+** teams.

POSITIONS OF RESPONSIBILITY

Professional Shows, Saarang | Deputy Coordinator   *Jun 2023 – Apr 2024*
- Managed the main event with **80k+ footfall** including artists like **Farhan Akhtar and Matisse & Sadko**.
- Spearheaded the publicity and sales for the online degree section of IIT Madras, comprising of **68K+ students**.

Corporate Relations: E-Cell | Assistant Manager   *Dec 2022 – Mar 2023*
- Solely responsible for negotiating a deal-in-kind worth **INR 2.5L** from **Kokuyo Camlin Ltd.** for the E-Summit.
- Managed the **branding and advertisement** of sponsors during E-Summit and **chaperoned VIP delegates**.

AI Club | Deputy Coordinator   *Dec 2022 – Mar 2023*
- Created an **LSTM** in PyTorch, trained it on **CNN11/DailyMail dataset** to summarize text in <**100 words**