Sampad Kumar Kar

Email: sampaddavcspur2001@gmail.com
sampadk04@cmi.ac.in
Portfolio: sampadk04.github.io
Mobile: +91-797-8620-354
Github: github.com/sampadk04

Education

- **Chennai Mathematical Institute** — Chennai, India
  M.Sc. - Computer Science; ***GPA: 8.94****; transcript* — *2022 - 2024*
- **Chennai Mathematical Institute** — Chennai, India
  B.Sc.(Hons.) - Mathematics and Computer Science; ***GPA: 8.82****; transcript* — *2019 - 2022*

Experience

- **Michelin** — Pune (Hybrid), India
  Data Science Intern — *Dec 2023 - May 2024*
  - **Projects**: TireScan
  - **Contribution**: Developed TireScan, a system for extracting tire information from web data using LLMs. Implemented and compared four algorithms, enhancing accuracy and efficiency. Documented project methodology and results, highlighting the RAG pipeline for its high accuracy. *Report*
- **BNY Mellon** — Chennai, India
  Financial Modelling Intern — *May 2023 - July 2023*
  - **Project**: Enhancing Short Term Cash Flow Model
  - **Contribution**: Improved BNY Mellon's STCF forecasting by introducing quantitative methods for identifying seasonal patterns, enhancing accuracy across LOBs and currencies. Conducted statistical tests and modeling to validate the quantitative framework, which was subsequently submitted to Risk for audit and adoption. *Report*
- **Trumpf** — Chennai, India
  Computer Vision Intern — *Nov 2022 - Apr 2023*
  - **Project**: Sheet Metal Part Identification
  - **Contribution**: Developed a deployable pipeline for sheet metal part identification, using transfer learning on pre-trained models to classify images into 10 predefined categories and output the corresponding CAD files. *Report*
- **Launchpad.ai** — Remote
  Machine Learning Intern — *July 2022 - Sep 2022*
  - **Project**: Fitness Activity Recognition (in collaboration with Nike)
  - **Contribution**: Developed a pipeline for comparing trainee and expert fitness videos. Utilized 'Dynamic Time Warping' on coordinates extracted via 'Mediapipe Blazepose' to handle variations in camera angles. *Report*

Notable Projects

- **Clustering - Fast Text based Clustering**: Developed four clustering techniques, including two optimized versions of sklearn's KMeans and two novel techniques designed from scratch to efficiently cluster large datasets using 'Jaccard Similarity' as the metric. *GitHub* (April '22)
- **Computer Vision - Facial Emotion Detector**: Created a real-time facial expression detector using Deep Residual Networks upon the *Kaggle Facial Expression Dataset*, with deployment for live webcam-based emotion recognition. *GitHub* (Jan '23)

Notable Reports

- **Master's Thesis (LLM) - SSMs: An Efficient Alternative to Attention**: This thesis investigates state space models (SSMs) as alternatives to Transformers for deep sequence modeling. It examines the efficiency and performance of SSM variants, including S4, Mamba, LRUs, and Griffin, through experiments on text, image, and audio generation tasks using tiny language models. *GitHub Report* (May '24)
- **ANN - Comparison of Regularization Techniques in DNNs**: (Reading Project): Comparative research on several popular regularization techniques using real world weather dataset. Also tested this on their own test set, to validate the conclusions based on training and validation errors, to come up with the best regularization paradigm. *Report* (Jan '22)

TA Experience

- **Chennai Mathematical Institute** — Chennai, India
  Teaching Assistant — *Jan 2023 - Apr 2024*
  - **Data Mining and Machine Learning**: Instructors: Prof. Madhavan Mukund, Prof. Pranabendu Misra
  - **Advanced Machine Learning**: Instructors: Prof. Pranabendu Misra
  - **Natural Language Processing**: Instructors: Prof. Ramaseshan Ramachandran

Honors and Awards

- Shriram Scholarship (Full Tution Fee Waiver and Stipend at CMI for Masters) - 2022 - 2024
- Shriram Scholarship (Full Tution Fee Waiver and Stipend at CMI for Bachelors) - 2019 - 2022
- NTSE Scholar, INMO Merit Awardee - 2017
- KVPY, RMO, NSEP, NSEA Qualifier - 2017-18
- JEE Advanced AIR **4062**, JEE Mains AIR **3042** - 2019

Technical Skills

- **Languages**: Python, R, Matlab C++, Java, Haskell
- **ML**: Scikit, PyTorch, transformers, LangChain, NLTK, OpenCV
- **Cloud**: Azure, AWS, VertexAI