

Arun Pa Thiagarajan

<https://www.arunppsg.in/>

Email : arunppsg@gmail.com

Github: [arunppsg](#)

EDUCATION

- **PSG College of Technology**

Integrated Masters in Data Science; GPA: 3.63 (9.08/10.0)

Coimbatore, India

July 2016 – May 2021

PUBLICATIONS, POSTERS AND PATENT

- Advika Vidhyadhiraja, Arun Pa Thiagarajan et. al, *Open Source Infrastructure for Differentiable Density Functional Theory* (workshop paper) In: SynS & ML Workshop @ ICML 2023.
- Bharath Ramsundar and Arun Pa Thiagarajan, *Chiron: A Cloud Scientific Machine Learning Programming Environment*, 2022 (patent), United States Provisional Patent Application.
- Arun Pa Thiagarajan, *Potential Biases in Using Machine Learning for Healthcare Applications* (poster). In: RBCDSAI-FCAI Conference on Deployable AI, 2022
- Arun Pa Thiagarajan et. al *Data-Driven Analysis of Food Corporation of India's Operations and Policy Recommendations*, 2021 (report).

EXPERIENCE

- **Deep Forest Sciences**

Machine Learning Engineer (remote)

Fremont, CA

Mar '22 - Present

- Implemented, performed distributed training and benchmarking of graph neural networks and other deep learning algorithms over molecular datasets.
- Designed and built a machine learning platform for computational drug discovery applications on top of AWS services.
- The platform is powered by a web-interface built using React framework, a backend using RestAPIs for processing user requests, a MySQL database to store user data and AWS Batch Service to handle machine learning jobs like model training and inference.

- **Indian Institute of Technology, Madras**

Project Assistant

Chennai, India

July '21 - Feb '22

- Built a malware detection framework and implemented predictive modelling algorithms for detecting malware from network traffic data. The framework incorporated machine learning algorithm for classifying flows in a network, algorithms to detect suspicious domain names based on DNS patterns and TLS fingerprinting technique to detect malicious servers.
- Built an open-source packet logging tool which can capture and parse network packets at high speed (tested upto 50 MB/sec) using ring buffer and memory mapping techniques to capture and parse incoming packet headers and detect duplicate packets using a BloomFilter(code).

- **Tata Consultancy Services**

Research and Development Intern

Chennai, India

May '19 - Nov '19

- Analyzed time series sales data and built time-series forecasting models for prediction future sales.
- Studied dynamic pricing of interdependent and perishable products using reinforcement learning techniques and proposed a Q-Learning model which achieved a yield of 16% higher than passive pricing techniques.

OPEN SOURCE CONTRIBUTIONS

- Contributed machine learning model architectures to the open scientific ML library DeepChem and improved testing infrastructure by contributing to CI/CD pipeline and increasing test coverage (pull requests).
- Implemented the paper Time Series Anomaly Detection using Generative Adversarial Networks for performing anomaly detection in time series data (code).
- Contributed patches to PyTorch (contributions), a popular deep learning framework and PyTorch-Geometric, a deep learning framework for implementing graph neural networks (contributions).

TALKS

- Gave a lightning talk at Pycon India 2021 about DeepChem and a 20-min talk at December 2022 monthly FOSS United Bangalore meetup.
- Gave a talk on Technical Documentation in Regional Languages at FOSS Goa, 2023 meetup

OTHERS

- Completed the course on Deep Generative Models offered by Center of Continuing Education, Indian Institute of Science, Bangalore in May 2023.
- Awarded Achievement Award for outstanding curricular, co-curricular and extra-curricular achievements in MSc Data Science class of 2021.
- Attended ACM Winter School on Cybersecurity held in Dec 2019 at NISER, Bhubaneswar
- Conducted *Cricket and Statistics* program at Mango Education for kids aged between 11 - 15.
- Volunteer at FOSS United - a non-profit organisation which promotes open source software ecosystem in India.

PROGRAMMING SKILLS

- **Programming Languages:** Python, C, C++
- **Technologies and Frameworks:** MongoDB, MySQL, PyTorch, Git, Github Actions, Docker, AWS Cloud Infrastructure

Last updated: October 10, 2023