

Arun Pa Thiagarajan

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

Email : arunppsg@gmail.com

Github: [arunppsg](#)

EDUCATION

- **PSG College of Technology**

5 Years 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, 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 al *Data-Driven Analysis of Food Corporation of India's Operations and Policy Recommendations*, 2021 (unpublished work, [report](#)).

PROFESSIONAL EXPERIENCE

- **Deep Forest Sciences**

Machine Learning Software Engineer (remote)

Fremont, CA

Mar '22 - Present

- Compared various model architectures and pre-training strategies for molecular representation learning by evaluating their performance through benchmarking on downstream molecular property prediction tasks.
- Preliminary results from the study revealed that Graph Transformer based architectures out-performs BERT-like transformer architectures in 3 out of 5 downstream prediction tasks.
- Design and led the development of a web-based platform for computational drug discovery. The platform uses React, REST APIs, MySQL, AWS services, and Git to manage user interactions, data storage, machine learning tasks and version control.
- Created CI/CD pipelines for testing and deployment via docker images of the application and maintained infrastructure as code using terraform to support the computational needs for the organization.
- Worked with VPCs, DNS configuration, TLS, nginx, and load balancers to configure and manage user requests for the platform securely.

- **Indian Institute of Technology, Madras**

Project Assistant

Chennai, India

Jan '21 - Feb '22

- Developed a malware detection framework for detecting malware from network traffic data. The framework integrated machine learning algorithms for classifying network flows, algorithms for detecting suspicious domain names based on DNS patterns, and TLS fingerprinting techniques to identify malicious servers.
- Created an open-source packet logging tool capable of capturing and parsing incoming network packets at high speeds (tested up to 50 MB/sec) utilizing ring buffer and memory map ([code](#)).

- **in-d.ai**

Machine Learning Engineering Intern

Chennai, India

June '20 - August '20

- Developed a convolutional neural network based model using connectionist temporal classification loss for detecting handwritten characters and symbols from documents.
- The model achieved an accuracy of 74% on test data and it was deployed using a Flask-based web framework for text extraction from documents.

- **Tata Consultancy Services**

Research and Development Intern

Chennai, India

May '19 - Nov '19

- Investigated dynamic pricing strategies in e-commerce sector using reinforcement learning techniques and developed a Q-Learning model that outperformed static pricing techniques by 11% in terms of revenue for the seller.

OPEN SOURCE CONTRIBUTIONS

- Contributed machine learning model architectures to the open-source scientific machine learning library [DeepChem](#) and enhanced its testing infrastructure by contributing to the CI/CD pipeline and expanding test coverage ([pull requests](#)).
- Implemented a PyTorch version of the paper [Time Series Anomaly Detection using Generative Adversarial Networks](#) for performing anomaly detection in time series data ([code](#)).
- Contributed code patches to PyTorch ([contributions](#)), a widely used deep learning framework, and PyTorch-Geometric, a deep learning framework specifically designed for implementing graph neural networks ([contributions](#)).

TALKS

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

PROGRAMMING SKILLS

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

OTHERS

- Earned an A+ grade in the Deep Generative Models online course offered by the Center of Continuing Education at the Indian Institute of Science, Bangalore, in May 2023 ([certificate](#)).
- 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.
- [Book Notes](#) - detailed notes from the books I have read.
- [Technical Posts](#) - a list of my technical posts on topics related to software engineer, Python and machine learning.

Last updated: March 14, 2024