Arun Pa Thiagarajan

https://www.arunppsg.in/

Email : arunp.psg@gmail.com
Github: arunppsg

Education

- **PSG College of Technology** — Coimbatore, India
  5 Years Integrated Masters in Data Science; GPA: 9.08/10.0 — *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 al *Data-Driven Analysis of Food Corporation of India's Operations and Policy Recommendations*, 2021 (unpublished work, report).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** — Bangalore, India
  Machine Learning Software Engineer (remote) — *Mar 22 - Present*
  - Design and led the development of a web-based platform for computational drug discovery. The platform uses React, RestAPIs, 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 orchestrated infrastructure using Terraform to support the computational needs for the organization.
  - Worked on networks, security groups, DNS configuration, nginx, and load balancers to configure and manage user requests for the platform securely.
  - Compared various model architectures and pre-training strategies for molecular representation learning by evaluating the learned embeddings on downstream molecular property prediction tasks.

- **Indian Institute of Technology, Madras** — Chennai, India
  Project Assistant — *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.

- **in-d.ai** — Chennai, India
  Machine Learning Engineering Intern — *June '20 - August '20*
  - Built a document classification system using a linear classifier on TF-IDF representation of the documents.
  - Developed a convolutional neural network based model using connectionist temporal classification loss for detecting handwritten characters and symbols from documents.

- **Tata Consultancy Services** — Chennai, India
  Research and Development Intern — *May '19 - Nov '19*
  - Analyzed time-series sales data and built machine learning pipelines using time-series forecasting models for predicting future sales.
  - 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

- Introduced state-of-the-art machine learning model architectures to the open-source scientific python library DeepChem, introduced new APIs, enhanced testing infrastructure and created tutorial notebooks for users (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).
- Created an open-source TCP/IP 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).

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, Bhubaneshwar.
- 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: June 21, 2024