Arun Pa Thiagarajan

https://www.arunppsg.in/

Email : arunppsg@gmail.com

Github: arunppsg

Education

- **PSG College of Technology** — Coimbatore, India
  Integrated Masters in Data Science; GPA: 3.63 (9.08/10.0) — *July 2016 – May 2021*

Publications, Posters and Patent

- Advika Vidhyadhiraja, **Arun Pa Thiagarajan** et. al (2023), *Open Source Infrastructure for Differentiable Density Functional Theory*. In: SynS & ML Workshop @ ICML 2023.
- 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 (2021) Data-Driven Analysis of Food Corporation of India's Operations and Policy Recommendations (report).
- Chiron: A Cloud Scientific Machine Learning Programming Environment, United States Provisional Patent Application, Bharath Ramsundar and Arun Thiagarajan.

Experience

- **Deep Forest Sciences** — Fremont, CA
  Machine Learning Engineer (remote) — *Mar '22 - Present*
  - Implemented, trained and benchmarked graph neural networks and various other kind of deep learning networks over large datasets on multi-gpu multi-node platform with applications in molecular property prediction.
  - Designed and built a maching learning platform for computational drug discovery applications on AWS Batch service.
  - Built backend RestAPIs using FastAPI webframework and MySQL database for submitting jobs to the platform and developed backend services.

- **Indian Institute of Technology, Madras** — Chennai, India
  Project Assistant — *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** — Chennai, India
  Research and Development Intern — *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 anomalye detection in time series data (code).
- Contributed patches to PyTorch (contributions), a popular deep learning framework and PyTorch-Geometric, a popular deep learning framework for implementing graph neural networks (contributions).

Others

- Awarded Achievement Award for outstanding curricular, co-curricular and extra-curricular achievements in MSc Data Science class of 2021.
- Conducted *Cricket and Statistics* program at Mango Education for kids aged between 11 - 15.
- Budget Analyzer: An application to keep track of income and expenses and predicting future expenses using time series analysis. Tools and technologies used: MongoDb for backend and PyQt for front end. (code).
- Took Deep Generative Model courses from the Center of Continuing Education, Indian Institute of Science, Bangalore in May 2023.
- 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 2, 2023