Ananth Mahadevan

Phone: +358417170319
Email: ananth1996@gmail.com

GitHub
Website
LinkedIn

Profile

PhD candidate with extensive research experience in algorithmic data science and machine learning. Proven expertise in designing scalable data pipelines, with hands-on experience at leading institutions and organizations. Published author with a track record of impactful contributions to machine learning research and industry applications.

Skills

Programming Languages: Python, C/C++, Shell, R

Machine Learning: PyTorch, TensorFlow, JAX, CUDA, scikit-learn, NumPy, Pandas

Graph and Network Tools: NetworkX, PyTorch Geometric, SNAP

Database and Big Data: Apache Spark, Neo4J, MariaDB, Elasticsearch, Dagster

Miscellaneous: Excellent troubleshooting and debugging skills

Education

- **University of Helsinki** — Helsinki, Finland
  PhD in Computer Science — *Aug 2020 - present*
  - Doctoral Researcher in Algorithmic Data Science group
  - Supervisor: Prof. Michael Mathioudakis
  - **Thesis Topic**: Scaling and Maintaining Machine Learning Pipelines

- **Aalto University** — Otaniemi, Finland
  Master of Science in Machine Learning, Data Science and AI — *Sep 2018 - May 2020*
  - Full Tuition Scholarship Student
  - **Thesis Topic**: Vote Prediction Models for Signed Social Network

- **R.V College Of Engineering** — Bangalore, India
  Bachelor of Engineering in Computer Science and Engineering — *Aug 2014 - May 2018*
  - 3rd rank in department
  - **Thesis Topic**: Extract-Transform-Load of Financial Transaction Tax data using Elasticsearch

Work Experience

- **Project Researcher** — Helsinki, Finland
  University of Helsinki — *May 2022 - present*
  - High Performance Computing for the detection and analysis of Historical Discourses (HPC-HD)
  - Collaboration between Helsinki University, Aalto University and Turku University
  - Implemented an **end-to-end data science pipeline** for analyzing millions of digitized historical documents
  - Utilized Apache Spark to scale up to $6\times$ more raw data, and Dagster for workflow orchestration

- **Doctoral Researcher** — Helsinki, Finland
  University of Helsinki — *Aug 2020- present*
  - Conducted research on Scaling and Maintaining Machine Learning Pipelines

- **Teaching Assistant** — Helsinki, Finland
  University of Helsinki — *Jan 2021- Mar 2021*
  - Worked with *Prof.* Michael Mathioudakis for the Network Analysis Course
  - Created and graded five assignments
  - Conducted four tutorials on advanced graph learning concepts

- **Research Assistant**   Otaniemi, Finland
  Aalto University   *May 2019- May 2020*
    - Worked with *Prof.* Aristides Gionis and Bruno Ordozgoiti in the Data Mining Group
    - **Wikipedia RfA Project**: predicting voter behaviour in Wikipedia elections networks
    - **Diversified Item Selection Project**: implemented and scaled up experiments

- **Teaching Assistant**   Otaniemi, Finland
  Aalto University   *Sep 2019- Dec 2019*
    - Worked with *Prof.* Aristides Gionis on the Algorithmic Methods of Data Mining course
    - Created three assignments and arranged exercise sessions
    - Managed and graded the course project on balanced graph clustering

- **Summer Thesis Worker**   Bangalore, India
  Goldman Sachs   *Jan 2018 - May 2018*
    - Worked in the **Client Experience Engineering** team on **client email prioritization**
    - Used **Conditional Random Fields** (CRFs) to tag and remove *disclaimers* from client emails
    - Performed **sentiment analysis** and Named Entity Recognition(NER) on emails
    - Platforms Utilized: pycrfsuite, sklearn, Protobuf, spaCy, Stanford CoreNLP

- **Summer Analyst**   Bangalore, India
  Goldman Sachs   *May 2017 - July 2017*
    - Worked in the **Tax Operations Technology** team
    - Imported large volume tax from DB2, Sybase IQ databases to **Elasticsearch** cluster
    - Set up, maintained and optimized multi-node Elasticsearch cluster in production environment

Hobbies and Extracurricular Activities

Parliamentary Debating
- **President** of Helsinki Debating Society (HDS) : Sep 2020 – Jan 2023
- **Finnish Debate Champion 2019** at FINDA University Championship 2019, Finland

Publications

- Ananth Mahadevan and Michael Mathioudakis. Cost-aware retraining for machine learning. *Knowledge-Based Systems*, 293:111610, 2024

- Ananth Mahadevan, Michael Mathioudakis, Eetu Mäkelä, and Mikko Tolonen. Optimizing a data science system for text reuse analysis. *arXiv*, 2024

- Yann Ryan, Ananth Mahadevan, and Mikko Tolonen. A comparative text similarity analysis of the works of bernard mandeville. *Digital Enlightenment Studies*, 1:28 – 58, 2023

- David Rosson, Eetu Mäkelä, Ville Vaara, Ananth Mahadevan, Yann Ryan, and Mikko Tolonen. Reception reader: Exploring text reuse in early modern british publications. *Journal of Open Humanities Data*, Apr 2023

- Ananth Mahadevan, Arpit Merchant, Yanhao Wang, and Michael Mathioudakis. Robustness of sketched linear classifiers to adversarial attacks. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, pages 4319–4323, 2022

- Arpit Merchant, Ananth Mahadevan, and Michael Mathioudakis. Scalably using node attributes and graph structure for node classification. *Entropy*, 24(7), 2022

- Ananth Mahadevan and Michael Mathioudakis. Certifiable unlearning pipelines for logistic regression: An experimental study. *Machine Learning and Knowledge Extraction*, 4(3):591–620, 2022

- Bruno Ordozgoiti, Ananth Mahadevan, Antonis Matakos, and Aristides Gionis. Provable randomized rounding for minimum-similarity diversification. *Data Mining and Knowledge Discovery*, 36(2):709–738, March 2022