

HARDIK PRABHU — CURRICULUM VITAE

Machine Learning Researcher (Remote) – CloudAEye, Inc.

📞 (+91) 9420726940 • ✉ hardik.prabhu@gmail.com

🌐 hardikprabhu.github.io • in hardik-prabhu • 🌱 HardikPrabhu

HIGHLIGHTS

- 3+ years of experience in machine learning research and development.
- 6 first-author publications in high-impact venues. (AAAI, ACM, Nature Scientific Reports)
- Academic and research experience at tier 1 institutes in India. (CMI, IISc)
- Machine Learning Researcher/Engineer experience in a tech startup.

EXPERIENCE

CloudAEye, Inc.

Fremont, CA, USA (remote)

Machine Learning Researcher

July 2024 - Present

- Currently developing an inference-level Retrieval Augmented Generation system, integrating LLMs (both open-source and proprietary), to address repository-level tasks such as automated code review
- Enhanced and optimized sequential log anomaly detection algorithm

Indian Institute of Science (IISc)

Bengaluru, India

Research Associate

Nov 2023 - June 2024

- Led a project on applying a novel GANs-based solution in energy anomaly detection.
- Authored a paper on the findings as first author, which was presented at the AI4TS (Artificial Intelligence for Time Series) workshop at the AAAI 24 (Association for the Advancement of Artificial Intelligence) conference.

FLAME University

Pune, India

Research Associate

Jan 2023 - Nov 2023

- Led projects focusing on multiple areas including Explainable AI and Genetic Algorithms.
- Served as a Teaching Assistant for courses related to Computational Modelling, Quantitative Methods and Machine Learning.
- Produced 2 Q1 journal papers, an ACM BuildSys workshop paper, and a Springer book chapter as first author.

CloudAEye, Inc.

Fremont, CA, USA (remote)

Machine Learning Engineer

July 2021 - Oct 2022

- Developed and deployed advanced deep learning and machine learning solutions specializing in anomaly detection within logs and metrics produced by cloud-native applications.
- Utilized deep learning techniques such as LSTMs, Variational Autoencoders (VAEs) and Normalizing Flows.
- Conducted technical interviews of candidates applying for the ML Engineer role.

CMI Algolabs

Chennai, India

Research Intern

May 2020 - Aug 2020

- Created a Python-based tool for a software company for mapping functionality script to software documentation by applying Latent Dirichlet Allocation, a topic modelling technique.

EDUCATION

Chennai Mathematical Institute (CMI)

Chennai, India

Master of Science in Data Science, CGPA : 8.38/10

Aug 2019 - May 2021

D.G Ruparel College, Mumbai University

Mumbai, India

Bachelor of Science in Mathematics, CGPA : 8.75/10

Aug 2016 - April 2019

FIRST AUTHOR PUBLICATIONS

- [1] **Prabhu, H.**, Valadi, J.K. and Arjunan, P., Generative Adversarial Network with Soft-Dynamic TimeWarping and Parallel Reconstruction for Energy Time Series Anomaly Detection. In Proceedings of AI4TS workshop of AAI, 2024 (**CORE Rank: A***)
- [2] **Prabhu, H.**, Valadi, J.K. and Arjunan, P., Explainable AI for Energy Prediction and Anomaly Detection in Smart Energy Buildings. In Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. (pp. 472-475). (**CORE Rank: A**)
- [3] **Prabhu, H.** and Arjunan, P., 2022, November. eptk: energy prediction toolkit. In Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. (pp. 512-515). (**CORE Rank: A**)
- [4] **Prabhu, H.**, Bhosale, H., Sane, A., Dhadwal, R., Ramakrishnan, V., and Valadi, J. 2024. Protein Feature Engineering Framework for AMPylation Site Prediction. Nature Scientific Reports. (**SCI IF: 4.6, Q1**)
- [5] **Prabhu, H.**, Sane, A., Dhadwal, R., Parlikkad, N.R. and Valadi, J.K., 2023. Interpretation of Drop Size Predictions from a Random Forest Model Using Local Interpretable Model-Agnostic Explanations (LIME) in a Rotating Disc Contactor. Industrial & Engineering Chemistry Research. (**SCI IF: 4.326, Q1**)
- [6] **Prabhu, H.**, Sane, A., Dhadwal, R., Siarry, P., Valadi, J. (2024). Metaheuristic and Evolutionary Algorithms in Explainable Artificial Intelligence. In: Valadi, J., Singh, K.P., Ojha, M., Siarry, P. (eds) Advanced Machine Learning with Evolutionary and Metaheuristic Techniques. Computational Intelligence Methods and Applications. Springer, Singapore.

NOTABLE PROJECTS

LocalRepoGPT at CloudAEye, Inc.

Environment: Python, LangChain, ChromaDB

2024 - Present

Currently developing an inference-level Retrieval Augmented Generation (RAG) system for automated code review and repository-level tasks. Key aspects:

- Options for integrating Large Language Models (LLMs), both open-source and proprietary.
- Implementing efficient retrieval mechanisms using vector databases.
- Designing systems for automated code review and repository analysis.
- Ensuring all operations are performed locally to address IP concerns and maintain data privacy.

Energy Time Series Anomaly Detection Project

Environment: Python, PyTorch

2023 - 2024

Developed a novel approach for detecting anomalies in energy time series data using Generative Adversarial Networks (GANs) with Soft-Dynamic Time Warping and Parallel Reconstruction. Key aspects:

- Implemented a DCGAN architecture with both WGAN and vanilla GAN training options
- Utilized Soft-Dynamic Time Warping for improved reconstruction loss
- Developed a parallel reconstruction method for efficient anomaly detection
- Benchmarked the model against multiple baseline methodologies on a dataset of 200 buildings
- First author paper accepted at the AI4TS Workshop @ AAI 2024

Energy Prediction Toolkit (eptk): An Open-source Python Package

Environment: Python

2021

Developed an open-source Python package for implementing and benchmarking energy-use prediction models on large datasets using standard performance metrics. Key aspects:

- Provides methods for engineering additional features (temporal, weather, and rolling stats) from datasets
- Implements ensembling techniques such as meta-regressors, Bayesian optimization, and subsampling to combine multiple models
- Developed a custom cross-validator for benchmarking models on time-series data
- First author paper accepted at BALANCES Workshop @ ACM BuildSys 2022

CERTIFICATION AND SKILLS

Programming: Python (PyTorch, TensorFlow, Scikit-learn, NumPy, Pandas, Pymoo, LangChain, SHAP)

Certifications: Teaching Certificate (FLAME University, 2023)

Relevant Coursework: Advanced Machine Learning, Bayesian Data Analysis, Multivariate Statistics, Reinforcement Learning, Natural Language Processing

Cloud and Other tools: AWS, Latex, and GitHub