HARDIK PRABHU — CURRICULUM VITAE

Research Associate
Centre for Mathematical Modelling, Centre for Computing and Data Sciences
FLAME University, Pune, India - 412115
(+91) 9420726940 • hardik.prabhu@gmail.com
linkedin.com/in/hardik-prabhu • https://github.com/HardikPrabhu
Blogs https://medium.com/@hardik.prabhu

RESEARCH INTERESTS

Generative Modeling, Deep Learning, Anomaly Detection, Explainable AI/Interpretable Machine Learning, Data Science and Applied ML.

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*

EXPERIENCE

FLAME University **Pune, India**
Research Associate *Jan 2023 - Present*
- Engaged in projects involving Explainable AI and Genetic Algorithms.
- Teaching Assistant for courses on Computational Modeling and Machine Learning.

CloudAEye, Inc. **Fremont, CA, USA (remote)**
Machine Learning Engineer *July 2021 - Oct 2022*
- Delivered sophisticated Deep learning/ Machine learning-based solutions for anomaly detection in both logs and metrics generated from cloud-native applications.
- Researched and developed a machine learning-based solution for the root cause analysis for cloud-native applications.
- Member of the recruitment panel. Responsible for conducting technical interviews of the candidates.

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.

RESEARCH PUBLICATIONS

Refereed Journal articles .....

[J.1] **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**)

Refereed Conference and Workshop papers

[C.1] **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**)

Accepted for publication

Book Chapters

[B.1] Siarry, P., Valadi, J.K., **Prabhu, H.**, Sane, A., & Dhadwal, R. (2024, expected). Metaheuristic and Evolutionary Algorithms in Explainable Artificial Intelligence. In J.K. Valdi, M. Ojha, K.P. Singh, & P. Siarry (Eds.), Advanced Machine Learning with Evolutionary and Metaheuristic Techniques (pp. XX-XX). Springer: Computational Intelligence Methods and Applications.

Posters and Demo Papers

[P.1] **Prabhu, H.**, Valadi, J.K. and Arjunan, P., (2023 expected). Explainable AI for Energy Prediction and Anomaly Detection in Smart Energy Buildings. Energise 2023: hosted by Alliance for an Energy-Efficient Economy (AEEE).

Publications Under Review

[C.2] **Prabhu, H.**, Valadi, J.K. and Arjunan, P., 2023, Exploring 1D Wasserstein DCGAN for Effective Anomaly Detection in Smart Building Energy Time Series Data. 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.(**CORE RANK : A**)

[J.2] **Prabhu, H.**, Ravishankar, C., Ganesan, A., Bhosale, H., Parlikkad, N.R, Siarry, P. and Valadi, J.K.,2023, Enhancing Random Forest Model Prediction of Gas Holdup in Internal Draft Air-Lift Loop Contactors: Leveraging Genetic Algorithms for Hyperparameter Tuning and SHAP for Interpretability. Engineering Applications of Artificial Intelligence. (**SCI IF : 7.802, Q1**)

RESEARCH AND DEVELOPMENT PROJECTS

Energy Prediction Toolkit: An Open-source Python Package

Environment: Python *May 2021 - Dec 2021*

Energy Prediction Toolkit (eptk) is a Python package for implementing and benchmarking energy use prediction models on a collection of large datasets using standard performance metrics. The package includes a variety of predictive models along with a set of configurations that were picked from the top performers in the ASHRAE - Great Energy Predictor III competition hosted on Kaggle. The package provides methods for engineering additional features (temporal, weather and rolling stats) from the datasets. The package also provides ensembling techniques such as meta-regressors, Bayesian optimization and subsampling to combine multiple models. A custom cross-validator is employed which is used for benchmarking models on Time-series data.

Doc2Script: Mapping Functionality Scripts to Software Documentation

Environment: Python *May 2020 - Aug 2020*
A Python-based solution, developed for a software company, employing Latent Dirichlet Allocation (LDA), a probabilistic topic modelling approach, to establish connections between functionality scripts and pertinent documentation available on the company website. This tool facilitates the precise mapping of scripts to corresponding HTML documentation, thereby optimizing the process of functionality testing.

TEACHING EXPERIENCE

- **Teaching Assistant** CSIT: 331, Machine Learning I (Jan 2023 - May 2023)
- **Teaching Assistant** CSIT: 121, Computational Modeling (Jan 2023 - May 2023)

PERSONAL PROJECTS

Application of Bayesian Optimization for Hyper-parameter tuning

Executed the application of Gaussian Process methodology to fine-tune the hyperparameters of a Convolutional Neural Network (CNN) that was specifically designed for the task of recognizing hand gestures.

Mathematical Foundations of PCA and a python implementation using numpy

Dimensionality reduction transforms data from high-dimensional spaces to lower dimensions, minimizing information loss. It serves purposes like noise reduction, data visualization, cluster analysis, and aiding further analyses. An illustrative technique is Principal Component Analysis (PCA), investigated in this project. The project involves a thorough exploration of PCA's mathematical foundation and its implementation from scratch using numpy.

Reinforcement Learning Algorithms to solve-Gridworld Problems

Trained an agent to travel a M x N grid from any arbitrary cell to the terminal cell, and avoiding obstacles placed in between by applying various RL algorithms such as Policy Iteration, Monte Carlo Simulations, SARSA, Q learning, Temporal Difference and Semi-gradient Approaches.

CERTIFICATION AND SKILLS

Custom Models, Layers, and Loss Functions with TensorFlow

Cerificate Issued by DeepLearning.AI *Jan 2022*

Custom and Distributed Training with TensorFlow

Cerificate Issued by DeepLearning.AI *August 2022*

Programming Languages: Python, R and LaTeX

Python Packages: Pytorch, Tensorflow, Scikit-learn, Numpy, Pandas, Pymoo and more.