

Dr. Tirthajyoti Sarkar

FREMONT, CA 94536

E-mail: tirthajyoti@gmail.com

Summary

Semiconductor technologist and data science manager with 18+ years of experience in R&D and product development. Currently driving innovation to apply data analytics/machine learning to semiconductor/manufacturing domain.

LinkedIn profile: <https://www.linkedin.com/in/tirthajyoti-sarkar-2127aa7/>

Data Science, Machine Learning articles: <https://medium.com/@tirthajyoti>

GitHub Homepage: <https://tirthajyoti.github.io>

Appointments

Position: Data Science and Solution Engineering Manager **January 2021 - Present**
Organization: Adapdix Corp.

Responsibilities: (a) Driving customer success for AI/ML solution development with semiconductor clients, (b) data science and ML solution architecture development for edge-based analytics, (c) researching novel solution ideas for expanding edge analytics business

Position: Senior Principal Engineer **December 2009 – January 2021**
Organization: ON Semiconductor, Sunnyvale, CA (formerly Fairchild)

Responsibilities: (a) AI/ machine-learning based yield improvement, (b) semiconductor design optimization framework development, (c) AI-based power IC, (d) Power Semiconductor technology and new product development (NPD) for applications in Automotive, Cloud infrastructure, gaming and AI, mobile, and industrial systems.

Position: Industry advisory member, Machine Learning lecturer (Pro-bono) **Aug 2019 - Present**
Organization: ValleyML.ai, Santa Clara, CA, IEEE/ACM

Responsibilities: (a) Conducting workshops on data science and machine learning, (b) AI/ML seminars, conference, and Expo organization

Position: Postdoctoral Research Associate and Research Assistant **August 2003 – November 2009**

Institution: University of Illinois at Chicago

Responsibilities: Worked on Federal agency-sponsored R&D projects (NSF, DOE, ONR)

Formal Education (Chronological Order)



Master of Science (MS) in Computational Data Analytics, Georgia Tech, College of Engineering. **August 2018 - now**



Doctor of Philosophy (Ph.D) in Electrical Engineering, University of Illinois at Chicago (UIC), Electrical and Computer Engineering. **March 2009**



Bachelor of Technology (B.Tech) in Instrumentation Engineering, Indian Institute of Technology (IIT), Kharagpur, India. **June 2003**

Continuing Professional Education (AI, Machine learning, Data science)



“Artificial Intelligence: An Introduction To Neural Networks And Deep Learning”, Stanford Univ.

July – August 2018



MIT Professional Program, “Data Science: Data to Insights”, Massachusetts Institute of Technology, Continuing Education.

May 2017 – July 2017

Core Competencies



✓ **Proven innovator:** Inventor/co-inventor on > 5 issued U.S. patents, multiple U.S. patents are in application stage.

✓ **Proven communicator of scientific study:** Author of 30+ international journal and conference papers in Tier-I category (IEEE or equivalent); **Senior Member, IEEE** (since 2015).

✓ More than **10,000 stars + forks** on GitHub open-source projects and repositories.

✓ 15+ years’ experience with numerical computing and finite-element simulations – modeling of semiconductors, switching electronics circuits, and electromagnetic components – Medici/T-Suprem/Ansys/Silvaco/PSPICE

✓ **Analytics in manufacturing, quality, and product development:** 10+ years of experience in analysis of semiconductor data - multivariate statistics, advanced graphing, control charts, hypothesis testing.

✓ **Software stack experience:**

- **GitHub public profile:** <https://github.com/tirthajyoti>
- **Programming Languages:** Python, R, Julia, JavaScript, Scala, VB.NET
- **Statistical modeling:** NumPy, Pandas, Scipy, Statsmodels, tidyverse, JMP (SAS), DataPower (TIBCO)
- **Scientific computation:** MATLAB/Simulink, Octave
- **Statistical Visualization:** Matplotlib, Seaborn, ggplot2, Tableau, Plotly, Bokeh, Altair, Vega-lite
- **Classical Machine Learning:** Scikit-Learn, Caret, kernlab, MLlib, XGBoost
- **Data wrangling, NLP, others:** Pandas, BeautifulSoup, SymPy, PySpark, SpaCy, NLTK, dplyr
- **Deep Learning:** TensorFlow, Keras, PyTorch
- **Distributed data processing:** PySpark, Ray, Dask, Databricks platform
- **Web technologies:** HTML5, CSS, JavaScript, D3.js, Node/NPM, Flask, Django
- **Cloud technologies (basic experience):** AWS (EC2, S3, RDS, DynamoDB, Sagemaker), Google Cloud

Books, Open-source packages, and data science/machine learning projects

Store ▾ Books Bundles Courses Featured Newsletters Podcast Support Why Leanpub

Statistics and Stochastic Simulations with Python

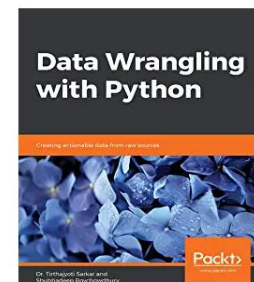
A Hands-on Approach

[Dr. Tirthajyoti Sarkar](#)

Python is the most wide-spread language for data science. But statistical modeling and stochastic simulation in Python are not well documented. This book tries to address that.

Dr. Tirthajyoti Sarkar

This book is 40% complete



BOOK: Productive and Efficient Data Science: Working on a book covering tools and techniques to perform data science tasks with high efficiency and productivity. Publisher: APRESS. Expected publication date: April 2022.

BOOK: Statistics and Stochastic Simulation with Python: Working on a book covering the essentials of statistical computing and stochastic simulation with Python. Expected publication date: January 2022.

BOOK: Data Wrangling with Python: Principal author of a book/courseware for [Packt publishing](#) on data wrangling techniques covering data scraping, cleaning, imputation, statistical plotting, etc.

Python Package: MLR: A lightweight Python package for doing statistical analysis on a regression problem - residual analysis and plotting, multicollinearity check, outlier detection, F-test, etc. [Here is the documentation.](#)

Python Package: Pydbgen: A lightweight Python library for generating random database tables. Useful for creating SQL database tables with synthetic data. [Here is the detailed documentation](#)

Python Package: UCI-ML API: An intuitive Python API to interface with the famous UC Irvine Machine Learning repository enabling an user to easily search and download relevant datasets or selectively choose a dataset based on its size or machine learning task categories. [Here is the detailed documentation.](#)

Python Package: DOEPY: Design of Experiment (DOE) is a critical activity for any scientist, engineer, or statistician planning to conduct scientific research. This is a wrapper library around the core packages (pyDOE and DiversiPy) to help generate various types of DOE matrices (random, Latin hypercube, face centered design, factorial matrix) from an arbitrary range of input variables. [Read the detailed documentation here.](#)

Selected professional experience/services

- Track Chair, *ValleyML AI Expo*, October 2021.
- Committee member, *AI and Robotics Expo*, Oct-Nov 2020.
- Content committee member, *Open Data Science Conference (ODSC), West*, 2020
- Data Science track chair, *State of AI/ML seminar*, 2019, 2020
- Editorial Associate, *Towards Data Science*, 2017-18
- Chair, Semiconductor Committee, *Power Supply Manufacturers' Association (PSMA)*
- Industry Expert member, Wide and Narrow bandgap technologies for Sustainable Energy Systems, IEEE
- Topic Chair, Sustainable Energy, *IEEE ECCE Conference*, 2016, 2017, 2018
- Technical Track Chairs in numerous other IEEE conferences.
- Visiting Lecturer, Indian Institute of Technology (IIT) Bombay, India, 2011-2012.



Selected Machine Learning projects (for semiconductor domain)

Semiconductor device design automation pipeline with machine learning

Goal: This project aims to automate/aid the complex technology development and device design tasks in the field of power semiconductors, using ML and optimization.

Tools/Techniques used: Scikit-learn, SciPy, JMP, R-part, nonlinear regressions, regularization, and various cross-validation strategies, decision trees, random forest, etc.

Deep learning-based semiconductor design feature extraction:

Goal: To use deep learning framework to mimic 'high-level' design experience of human experts by classification of designs into categories such as 'sub-optimal' or 'aggressive'.

Tools/Techniques used: TensorFlow, Keras, PIL, OpenCV, Scikit-image

Low power neural network for Power IC controller (US Patent filed)

Goal: Implement and embed deep learning function inside a controller IC for optimizing power conversion efficiency with a limited compute power and memory budget.

Tools/Techniques used: Keras/TensorFlow, SciPy.

Neural network-in-loop-SPICE modeling (US Patent filed)

Goal: neural network models and optimization modules for finding best semiconductor die design for a given electrical target using physically-scalable SPICE modeling data.

Tools/Techniques used: Scikit-learn, Keras/TensorFlow, SciPy.

Reinforcement learning based motor control algorithm and techniques (US Patent filed)

Goal: Build reinforcement learning based intelligent agent at local and supervisory level for distributed control of motor drives and optimize their robustness and performance dynamically.

Python-based data analytics framework for semiconductor manufacturing:

Goal: Build Python-based notebooks for analyzing silicon wafer data involving failure analysis, statistical plotting, normality analysis, and causality discovery by multivariate regression, outlier detection with Gaussian mixture models.

Tools/Techniques used: Scikit-learn, Statsmodels, Pandas, Seaborn.

Talks on data science/machine learning

- “GPU-powered Data Science and Analytics”, Blue Yonder AI Workshop, June 2021.
- “AI and Machine Learning for Power Electronics”, PSMA Webinar, April 2021
- “How to pursue a career in machine learning and AI”, Webinar, DuxVitae LLC, July 2020.
- “Data wrangling methods”, ValleyML.ai State of AI/ML, January 2020
- “Statistical methods for data science”, ValleyML.ai ML/DL boot camp, August 2019.
- “How AI/ML will drive emerging semiconductor markets”, Panel discussion, International SoC conference, UC Irvine, October 2019.
- “How AI and ML techniques can help ON Semiconductor business”, Panel discussion, Engineering Technology Forum, October 2019.