Tirna Deb, PhD

tirnadeb.astro@gmail.com — linkedin.com/in/tirna-deb-ph-d — github.com/drtirnadeb — https://drtirnadeb.github.io

Summary — Data-driven professional with 10+ yrs in data science, specializing in ML, large-scale analytics, & agile-driven development. Skilled in building scalable data pipelines, developing statistical models, and applying ML to extract insights from complex datasets. Led international teams across four continents, thriving in diverse environments, solving high-impact challenges. A fast learner, a relentless problem-solver, driven by curiosity and a passion for continuous learning. I push boundaries, embrace challenges, and see projects through to impactful outcomes. Now eager to apply my expertise in ML, predictive modeling, and big data analysis to drive data-driven decision-making & innovation in dynamic environments.

Technical Skills

- **Machine Learning algorithms & AI**: Predictive modeling, feature engineering, supervised & unsupervised learning, EDA, anomaly detection, time-series forecasting, customer behavior prediction, fraud detection.
- **Programming & Data Science**: Python (Pandas, NumPy, Scikit-learn, TensorFlow), Conda, SQL, Git.
- **Cloud & High-Performance Computing**: AWS, cloud computing (batch scripting, multi-GPU parallelization, HPC clusters).
- **MIT Applied Data Science Program (2024)**: 12-week AI, Python, and ML-focused program with real-world applications.
- **Soft Skills**: Leadership, management, mentoring, collaboration, creativity, innovation, analytical thinking, Agile.

Professional Experience

Harvard University, United States of America, *Rubicon Fellow* Nov 2023 – Present

- Led and managed critical data acquisition initiatives, securing ∼**$15M** in proprietary imaging data (30+ TB) for **30+ international stakeholders**, enabling large-scale statistical modeling and predictive analytics. Built and deployed a custom pattern recognition pipeline (Python), applying feature engineering and statistical clustering techniques to detect hidden structures in newly acquired datasets. Transformed raw observations into structured data products, engineered features, and predictive models, leading to a new classification framework that revealed previously unclassified patterns.

University of Western Cape, South Africa, *Postdoctoral Fellow* Apr 2022 – Oct 2023

- Developed an end-to-end data pipeline to process 30+ TB of radio telescope data, enabling high-precision signal detection. Applied **data preprocessing, noise filtering, and feature extraction using Python and distributed computing**, leading to the detection of new signals. **Automated workflows for anomaly detection** and interference removal, reducing processing time by **40%**. Created 50+ **data visualizations** using Python & VR-based tools, facilitating the detection, study, and classification of 100+ new objects.

University of Groningen, the Netherlands, *Graduate Researcher* Sep 2017 – Mar 2022

- Optimized a **3D source detection pipeline** using **threshold-based segmentation, region-growing algorithms, and statistical filtering** to extract faint signals from 15+ TB of noisy astronomical data. **Leveraged HPC and batch processing** for parameter tuning, boosting **detection rates by 200%**, resulting in the first-time detection of signals in 220 objects, and contributing to **one of the first discoveries from a state-of-the-art telescope**.
- To discover anomalies in a signal, built a model using information from a reference signal by **resampling data across three dimensions** and applying **double Gaussian fitting**. Constructed a **3D data representation** to analyze structural variations, leading to the **first-ever identification of neutral gas in a previously unexplored galactic system**.

Data Science Projects

- **Used Car Price Prediction:** Developed a **machine learning model** using **Random Forest, Decision Trees, and Linear Regression** to predict used car prices. **Engineered 16+ predictive features** and applied advanced encoding techniques, achieving $\mathbf{R^2 = 0.93}$ through hyperparameter tuning and cross-validation. Identified key price drivers (mileage, engine power, location), demonstrating how **data-driven pricing can enhance customer trust and market efficiency**.
- **Boston Housing Price Prediction:** Built a **linear regression model** to predict housing prices, optimizing **data preprocessing, feature engineering, and model evaluation** using **Python (pandas, NumPy, scikit-learn)**, achieving $\mathbf{R^2 = 0.76}$. Uncovered key pricing factors (crime rate, tax rates, proximity to employment centers), that will help real estate professionals **assess property values more accurately and refine investment decisions**.
- **FoodHub Order Analysis:** Analyzed customer order data to uncover demand patterns, pricing trends, and delivery inefficiencies using **Python (pandas, NumPy, Matplotlib)**. Recommended pricing optimizations and restaurant partnerships that **can increase sales and reduce delivery delays**.

Education

PhD in Astrophysics, University of Groningen, The Netherlands Sep 2017- Mar 2022
Bachelor & Master of Science in Physics, Presidency University, Kolkata, India July 2011-July 2016

- 15 peer-reviewed papers (7 as first author), 35+ talks at leading conferences around the world. Full list : Google Scholar.