

APARNA PRATURI.

Ph.D. Physicist & Data Scientist

India • +91-9177014526
aparnaps777@gmail.com
linkedin.com/in/aparna-praturi
github.com/Aparna-Praturi

Ph.D. physicist and Data Scientist specializing in mathematical modeling and ML pipelines. Expert in PySpark, scikit-learn, and MLOps, bridging advanced STEM research with production-ready AI. Actively expanding into Generative AI (RAG, Agentic AI) with a focus on rigorous, explainable systems.

PROFESSIONAL EXPERIENCE

Independent Data Scientist

Jan 2024 – Present

Remote / Self-Employed

- Built 5+ end-to-end ML solutions (forecasting, NLP) using PySpark, Docker, Airflow, and MLflow.
- Developed scalable preprocessing pipelines and interpretable models using SHAP, PCA, and t-SNE.
- Created business-focused dashboards in Power BI, translating complex analytics into actionable insights.

Postdoctoral Fellow

Mar 2022 – Dec 2023

Tata Institute of Fundamental Research (TIFR)

- Executed rigorous quantitative modeling using statistical regression on TMD datasets, leading to publication in *ACS J. Phys. Chem. C*.
- Designed efficient experimental workflows analogous to data annotation pipelines.
- Structured complex experimental datasets from precision optical measurements for downstream modeling.

Expert Answerer – Advanced Physics

Jan 2017 – Dec 2018

Chegg.com

- Authored detailed solutions for advanced mechanics and quantum mechanics, refining scientific communication skills.

SELECTED PROJECTS (2025)

CommerceSense

XGBoost • SARIMAX • Lift 1.73

- Forecasting:** Deployed models predicting chain-level weekly sales ($R^2 > 0.95$) and store-level sales (R^2 up to 0.75).
- Market-Basket Analysis:** Identified high-value product bundling opportunities, isolating association rules with high confidence and a maximum Lift of 1.73.
- Impact:** Identified high-lift, high-confidence bundles (Lift 1.73, Confidence 0.97) to increase average basket value and targeted cross-sell opportunities.

Voyant AI

Docker • Airflow • MLflow

- Engineered 3 scalable ML pipelines (Prediction, Recommendation, NLP) ensuring modular reproducibility.
- Integrated Flask APIs and Streamlit dashboards for real-time inference and experimentation/li>
- Impact:** Reduced manual experimentation and deployment effort by orchestrating one-click, scheduled pipelines, improving iteration speed and reliability.

MolecuLearn

SVM • SHAP • Optuna

- Developed interpretable ML pipeline classifying high-dimensional protein data (Weighted F1 0.68).
- Utilized PCA (95% variance) and Optuna for robust model selection; explained key signatures via SHAP.
- Impact:** Enabled domain scientists to trace model decisions back to biologically meaningful features, improving trust in predictions for exploratory analysis.

EDUCATION

MS Computer Science

Woolf University

2024 – 2026 (Exp)

Ph.D. Physics

IIT Bombay

2016 – 2022

Thesis: A study of metal nano-structures induced modification of light-emitter relaxation rates.

[View Certificate](#)

M.Sc. Physics

University of Hyderabad

2012 – 2015

[View Certificate](#)

TECHNICAL STACK

LANGUAGES & CORE

Python SQL MATLAB Mathematica

MACHINE LEARNING

Scikit-learn
Classification, Regression, Clustering PySpark
Forecasting NLP

MLOPS & TOOLS

Docker Kubernetes Airflow MLflow Git
PowerBI

PHYSICS/MATH

Stat Mech Quantum Mech
Numerical Methods

ACTIVELY LEARNING

RAG Agentic AI Prompt Eng

PUBLICATIONS

A. S. Praturi et al. *Relaxation and excitation rate modification: by metal nanostructures for solar energy conversion*. **J. Phys. Chem. C**, 125(15) (2021).

A. Praturi et al. *White-Light Spectral Interferometry for Characterizing Inhomogeneity*. **ACS Nanoscience Au**, 2(6) (2022).

J J Panda, K R Sahoo, **Aparna Praturi** et al. *High-sensitivity characterization of ultra-thin atomic layers using spin-Hall effect of light*. **J. Phys. Chem. C**, 125(15) (2021).

AWARDS & LEADERSHIP

AIR 25, CSIR–NET
Founder/Treasurer, Optica Student Chapter, IIT Bombay
IBM Data Science Specialization (Coursera)
Secretary, IITB Alumni Association Hyderabad