Abhijeet Anand, PhD

Postdoctoral Fellow Astrophysics & Data Science, Lawrence Berkeley National Lab, USA

abhijeetanand2011@gmail.com | +1 650-664-9558 | Milpitas, CA | LinkedIn | Scholar | GitHub

Astrophysicist and Data Scientist (PhD) with 5+ years building production-quality, scalable ML systems for very large datasets (60M+ records; 5+ TB/day). Expert in Python/pandas and statistical modeling (prediction, classification, hypothesis testing), with a track record of reducing false positives, improving accuracy, and deploying scientific libraries. Open-source contributor with 26 publications and hands-on software engineering experience.

Work Experience

Postdoctoral Fellow *Lawrence Berkeley National Lab, USA* *Sep 2022 - Present*

Data Engineering, ML and Big Data

- Built and deployed scalable **large-scale predictive modeling pipeline** (*redrock*, multi-class classification; PCA feature engineering) on 60M+ records, improving accuracy by 30% and cutting false positives by 30%.
- Designed **distributed real-time I/O pipelines (FITS, HDF5) to process 5+ TB/day, ≈100k rows/min** of structured and unstructured data, optimized for analytics in Python/pandas and SQL. Architecture is transferable to cloud-based systems (e.g., GCP, AWS).
- Developed and maintained **internal/external analysis libraries (open source)**, leading features, code reviews, and model improvements, and reporting workflows. Improved software quality via automated tests (unittest), documentation, and supporting reproducible analytics at scale.

Project Leadership and Scientific Impact

- Led **two cross-functional projects** with 15+ team members, coordinating software and deployment efforts to improve classification results for ongoing five-year survey phase.
- Proposed and implemented data-driven improvements to production pipelines, securing a successful **2-year project extension**. Recognized as a top 10% contributor within the collaboration.
- Mentored junior analysts/researchers on experimental design, statistical validation, and production-grade codeim-proving analysis quality and communication across teams.

PhD Research Fellow *Max Planck Institute for Astrophysics, Germany* *Sep 2018 - Jul 2022*

Data Engineering, ML and Big Data

- Built parallel **signal-processing models (matched-kernel detection)** for large sequential data, raising precision ($\geq 95\%$ purity) and reducing runtime from weeks to hours on 1M+ samples
- Developed and accelerated **non-linear regression pipelines (with Numba)** for parameter estimation, with rigorous model testing/validation and performance analysis.
- Ran large-scale statistical validation (hypothesis testing, bootstrapping) and experiment design to quantify uncertainty and build models.

Project Leadership and Scientific Impact

- Led **two large research projects**, from concept to publication, resulting in high-purity data products used by **50+ research teams worldwide.**
- Collaborated with computational physicists **to merge observational and simulated datasets**, enabling new insights into the physical origins of detected patterns.

Technical Skills

Programming	Python (NumPy, Pandas, SciPy, scikit-learn, Matplotlib), Git, LaTeX, Jupyter, unittest
ML & Statistics	Predictive modeling (regression/classification), PCA/feature engineering, clustering; statistics & probability (hypothesis testing, bootstrapping, experiment design).
Data Engineering	Parallel I/O pipelines (FITS, HDF5), HPC (Slurm, NERSC), Cloud-ready workflow design, Automated schedulers (cron)
Open Source	Maintainer/contributor: qsoabsfind, redrock (25+ GitHub stars), desispec (37 GitHub stars)
Soft Skills	Mentorship, Team leadership, Cross-functional collaboration, Agile workflows

Education

PhD in Astrophysics *Max Planck Institute for Astrophysics, Garching, Germany* *Sep 2018 - Jul 2022*

BS - MS in Physics *Indian Institute of Science (IISc), Bangalore, India* *Aug 2012 - Jun 2017*