Abhijeet Anand, PhD

Independent Fellow, Inter-University Centre for Astronomy and Astrophysics (IUCAA), Pune

abhijeetanand2011@gmail.com | +91 88615 59447 | Pune, India | LinkedIn | Scholar | GitHub

Data Scientist with a PhD in Astrophysics with 6+ years of experience building scalable ML and analytics solutions on large datasets (60M+ records; 5+ TB/day). Strong expertise in Python, SQL, statistical modeling, and predictive analytics. Proven track record of improving model accuracy, reducing false positives, and developing production-quality data pipelines. Experienced in cross-functional collaboration and open-source contributor with 30+ scientific publications and hands-on software development experience.

Work Experience

Independent Postdoctoral Fellow — *IUCAA, Pune* — *Dec 2025 - Present*
Affiliate Scientist — *Lawrence Berkeley National Lab, USA* — *Feb 2026 - Present*

- Working as an independent postdoctoral fellow, developing ML and data science methods for large scientific datasets and drafting scientific goals for upcoming national and international astronomical facilities.

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 controlled experiments to compare modeling approaches, analogous to A/B testing to guide data driven method selection. Also performed model evaluation metrics (ROC-AUC, precision-recall, etc).
- 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 performance and reliability, enabling more accurate data driven decisions in a large scale production pipeline.
- 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 codebases.
- Communicated analytical findings to both technical and non-technical audiences to support data driven decisions within a large international scientific collaboration.

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), SQL, Version control collaboration (Git workflows), LaTeX, Jupyter, unittest
ML & Statistics	Predictive modeling (regression/classification), NMF/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), debugging, dashboards, code optimization and documentation, Exploratory data analysis (EDA), visualization for insights and reporting
Open Source	Maintainer/contributor: qsoabsfind, redrock (27 GitHub stars), desispec (39 GitHub stars)
Soft Skills	Mentorship, Team leadership, Cross-team 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*