Boston, MA and Remote
(774) 212-6052
deividlsribeiro@gmail.com

Deivid Ribeiro, PhD

Data Scientist

GitHub
LinkedIn
Online Portfolio

Seeking a position where I can apply my passion for innovation, collaboration, technical excellence, and intellectual curiosity.

- Experienced in solving complex technical problems that require a deep understanding of large-scale software and hardware development, and in creating structure out of unstructured data.
- Expertise leading cross-functional initiatives to develop machine learning (ML) models and pipelines for deployment and continuous improvement in insights.
- Background in statistical modeling, time-series analysis, and probabilistic and causal methods, with expertise in building scalable ML infrastructure and high-throughput scientific analysis.
- Excellent communicator adept at explaining technical concepts and implications to diverse audiences.
- Strong leader, mentor, and project manager with a reputation for driving progress within cross-disciplinary teams.
- Education, research, and teaching experience from Columbia University and Brown University.

Skills

Programming Languages	C++, Python, Mathematica.
Tools & Platforms	Git, Linux, Docker, SQL, Machine Learning frameworks (TensorFlow, PyTorch).
Data Analysis	Statistical modeling, image processing, time-series analysis.
Languages	Fluent in English and Portuguese.

Professional Experience

Data Scientist (Remote) | University of Minnesota, Minneapolis, MN | Sept. 2022 to present

The Very Energetic Radiation Imaging Telescope Array System (VERITAS) is a ground-based gamma-ray observatory in Arizona, where four Cherenkov telescopes examine the universe's most violent and energetic processes.

Lead cross-functional initiatives in astronomy data analysis, software modernization, and machine learning integration, driving innovations across several major projects and objectives as the telescope nears retirement in the next three years. Transform technical capabilities through open-source adoption and reproducibility practices for scalable, interoperable analysis ecosystems.

Machine Learning

- Lead ML integration by training and deploying multiple ML models to improve data classification. Enhanced classification accuracy by implementing advanced feature extraction techniques.
- Conduct statistical analysis to validate model performance, including causal modeling and mediation analysis to inform model modification. Authored key code modifications that improved analysis sensitivity by 30% in a boosted decision-tree (BDT) model.
- Built and improved ML data pipelines and tools to automate processing and manage training and production layers.
- Developed algorithms and statistical models that use large-scale datasets to provide insights into physical systems.
- Created and applied analytical models using likelihood estimation and probabilistic reasoning to evaluate scenarios, predict outcomes, and identify potential roadblocks.

Software Engineering

- Modernized analysis stack by bridging legacy C++ systems with open-source python frameworks, enabling ML-driven workflows and full data model compatibility.
- Directed FAIR (findability, accessibility, interoperability, and reusability)-compliant modernization of 20-year raw data archive (>800 Tb) for systematic examination and data extraction.
- Led HPC infrastructure development for 1000+ concurrent projects across multiple institutions by creating a unified job submission framework with automated scheduling and error recovery.

Project Management & Leadership

- Chair analysis governance boards to assess statistical rigor and uncertainty in high-impact studies involving marginalized datasets. Lead expert review panels, clarify evaluation standards, and accelerate approval of non-standard techniques.
- Collaborate with colleagues on data modeling and reporting of performance metrics via presentations and publications.

- Revamped collaborative analysis and software development by engaging researchers, engineers, and stakeholders in planning and mentoring colleagues. Efforts also focused on integrating stalled features, attracting new contributors, establishing reproducible, modular pipelines, and streamlining code review processes.
- Mentored 5+ junior scientists in pipeline development and ML techniques, growing their expertise in Docker, Git, and high-performance computing workflows.
- Mentored grad students through validation testing, pipeline development, and codebase integration, culminating in successful PhD defenses

PhD Graduate Student | Columbia University | Sept. 2016 to Dec. 2022
Led design, development, and commissioning of a prototype next-generation telescope. Advanced data collection and analysis infrastructure for VERITAS, acknowledged with the Simon Swordy Outstanding VERITAS Contribution Award.

- Built and installed telescope's data-collection hardware. Oversaw hardware and software integration, including calibration and alignment tools.
- Enhanced real-time data processing pipelines to support rapid follow-up on alerts. Authored detailed usage guides for data analysis workflows.
- Enabled virtual observation during COVID by creating a VNC-based remote access framework, integrating secure connections across control systems, data acquisition nodes, and monitoring interfaces. Standardized VPN/VNC workflows to reduce onboarding time for internal users.
- Provided presentations and consultations to stakeholders on analytics results and solutions.

Teaching Experience and Academic Leadership Highlights

Columbia University
Research Mentor 2017 to 2022
Graduate Teaching Assistant Physics Department August 2016 - August 2018

- Recipient of the Allan M. Sachs Award for Outstanding Contributions to Teaching.
- Led 3 classrooms of about 20 undergraduate students in introductory physics lab course.

Physics Department Preceptor, Physics Department August 2017 - August 2018

- Developed and launched TA teaching training program, including a new peer teaching observation program workshop.
- Managed 50 graduate teaching assistants throughout the year.
- Designed and administered behavioral attitudes survey on introductory physics lab students.
- Reviewed and edited physics lab manuals.

Brown University
Science Center Fellow Summer 2013

- Produced a short film about Brown's Ladd Observatory.

Student Immigrant Movement (now known as Stories Inspiring Movement)
SIM is a grassroots organization that empowers immigrant youth and young adults in Massachusetts through storytelling, mentorship, relationship building, and organizing training.
Communications Director, Student Immigrant Movement May 2010 - August 2011

- Developed organization website, management system and media relations plan.
- Recruited and supervised student teams in community organizing, website management, and leadership training.
- Organized legislative hearing at Massachusetts State House on in-state tuition for undocumented immigrant students.

Education

Phd in Astrophysics, Columbia University - 2022
Bridge-to-PhD Scholar, Columbia University - 2016
Sc.B Physics, Brown University - 2013