

Zachary Hafen-Saavedra

PhD, Theoretical and Computational Astrophysics

z.hafen.saavedra@gmail.com || zhafen.github.io || Chicago, IL ||  

Summary

[\[click here for a work sample\]](#)

Computational scientist and analyst with ten years experience earned as a Northwestern University and UC Irvine astrophysicist. Extensive history leading interdisciplinary collaborations and communicating complex concepts. Seeking positions in data science, quantitative modeling, and software development.

Skills

Techniques: data analysis (time-series, cleaning, exploratory, big data, sparse), machine learning, natural language processing, frequentist/Bayesian statistics, code testing, nested/multiscale modeling

Interpersonal skills: technical leadership and management, storytelling, mentoring

Tools: Python (numpy, pandas, matplotlib, scikit-learn), SQL, TensorFlow, Unix, C/C++, git, nltk

Experience

McCue Prize Postdoctoral Fellow

University of California, Irvine

July 2020 - June 2023

Irvine, CA

- Employed [natural language processing](#) to convert >200,000 scientific abstracts to quantitative data
- Trained an ensemble voting model to use abstract word content and paper metadata to [predict citation count](#) to within 3 citations per year for 75% of the validation set
- Performed complex [filtering of >2 TB of remote data](#) via the NASA astrophysics data system API
- Spearheaded an international team spanning eight institutions to [validate Bayesian statistical models](#)
- Organized [a meeting of 20 leading galaxy experts](#) to build community consensus
- Improved [mock-data fidelity by >200%](#) by interfacing with open source atomic spectra data

GK-12 Graduate Fellow

Northwestern University

June 2014 - July 2020

Evanston, IL

- Created [quantitative visualizations](#), educational [explanatory visualizations](#), and award-winning [artistic visualizations](#) to communicate core messages
- Processed [tens of TB of >20-dimensional data](#) using high-performance-computing resources, reducing to <100 GB of highly-interpretable data
- Employed [modern software development best practices](#) (unit testing, version control, etc.) to maintain a broad suite of scientific software
- Utilized and upgraded a C code to generate [>100,000-CPU-hour simulations](#) of entire galaxies
- Performed [time-series decision-tree classification](#) to predict the extragalactic origin of Earth
- Developed [software for non-relational data management](#), including analysis of contained relational data
- Introduced [>100 students from underrepresented backgrounds](#) to data science by leading one of Chicago's first high-school data-science education initiatives
- Collaborated with a multidisciplinary range of scientists to publish (to date) [36 papers, 7 as a lead author](#)

Education

The Erdős Institute

Data Science Certificate

2023

Irvine, CA

Northwestern University

PhD, MS, Physics and Astronomy

Specialization: Theoretical and Computational Astrophysics

2020

Evanston, IL

University of Northern Colorado

BS, Mathematical Physics

2014

Greeley, CO