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. Exploring positions related to data science, data analysis, and quantitative modeling.

Skills

Techniques: data analysis (time-series, cleaning, exploratory, big data, sparse), machine learning, natural language processing, frequentist/Bayesian statistics, code testing, nested/multiscale modeling
Soft skills: technical leadership and management, public speaking, mentoring
Tools: Python (numpy, pandas, matplotlib, scikit-learn), SQL, TensorFlow, Unix, C/C++, git, nltk

Experience

McCue Prize Postdoctoral Fellow July 2020 - June 2020
University of California, Irvine Irvine, CA

- Trained an ensemble voting model to predict citation count to within 3 citations per year for 75
- Employed natural language processing to convert >200,000 scientific abstracts to quantitative data
- Performed complex filtering of $>$2 TB of remote data via the NASA astrophysics data system API
- Led an eight-institution, international collaboration to enable an analysis requiring expertise from observers, analysts, and simulators
- Forward-tested Bayesian statistical models against three increasingly-complex test cases
- Organized a meeting of 20 leading galaxy-shape experts to identify target measurements
- Improved mock-data fidelity by $>$200% by interfacing with open source atomic spectra data

GK-12 Graduate Fellow June 2014 - July 2020
Northwestern University 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 essential software
- Utilized and modified a C simulation code to generate >100,000-CPU-hour simulations of entire galaxies
- Developed software for non-relational data management, including analysis of contained relational data
- Performed time-series decision-tree classification to predict the extragalactic origin of Earth
- Brought opportunities to $>$100 students from underrepresented backgrounds by leading one of Chicago's first data-science education initiatives
- Collaborated with a multidisciplinary range of scientists to publish (to date) 36 papers, 7 as a lead author

Education

The Erdos Institute June 2023
Data Science Certificate Irvine, CA

Northwestern University May 2020
PhD, MS, Physics and Astronomy Evanston, IL
Specialization: Theoretical and Computational Astrophysics

University of Northern Colorado May 2014
BS, Physics, Math emphasis Greeley, CO