

Stefano Meschiari, Ph.D.



I am an experienced data scientist with a background in scientific research. I have worked on projects that involved researching new algorithms and creative approaches to attack hard problems; analyzing complex, label-poor datasets and extracting meaningful patterns to guide decision-making; building novel production ETL and ML platforms; helping to bring ML-powered product capabilities to fruition for users; and creating functional data tooling and APIs.

I enjoy working with cross-functional teams to create, iterate, and deliver from idea, to MVP, to production. In particular, I love collaborating with product and design teams to coherently integrate data science into their vision, and helping them iterate fast on new concepts and ideas with data-driven prototypes and customer discovery.

Recent Work Experience

SENIOR TECHNICAL LEAD
Oct 2021 - Dec 2021

TECHNICAL LEAD
May 2019 - Nov 2021

SENIOR DATA SCIENTIST
Jul 2017 - May 2019

PRODUCT DATA
SCIENTIST
Feb 2016 - Jun 2017

Duo Security (Cisco)

I led the research and technical development of the data science platform that powers the **Duo Trust Monitor** product feature all the way to release. My work at Duo over four years included:

- Developed pipeline components and infrastructure that analyze data, build models, and surface possible threats and authentication anomalies at scale, using Apache Spark, SparkML, and H2O.
- Researched foundational supervised and unsupervised algorithms tailored to the security domain, with a particular attention to simplicity, explainability, robustness, and scalability. Translated internal domain expertise into expert rule layers and heuristics. Analyzed our vast authentication dataset to mine new patterns of suspicious behavior.
- Collaborated with Product and Design teams to understand how to shape the algorithmic and presentation layers that would meet our product vision, simplify our customers' operations, and increase the value of the Duo platform. Worked with customers and domain experts via interviews and observations to test new capabilities and models.
- Created reports, dashboards, and prototypes to assist with product and engineering decisions and build consensus on technical and UX solutions.
- 2 patents submitted to USPTO through Cisco.

Civitas Learning

- Created and improved on prototype machine learning tools and pipelines to model university student outcomes.
- Prototyped and implemented new product ideas and internal tooling that employ machine learning, novel summary statistics and visualizations.
- Maintained and improved the custom modeling platform. Reduced batch training and scoring running time and cloud costs by half.

W.J. MCDONALD
POSTDOCTORAL
FELLOW
2012 - 2016

SAVE/POINT, PRINCIPAL
INVESTIGATOR
2014-2016

- Developed new machine learning models and custom classification algorithms (in R/Caret, SparkML, and JavaScript); back-end data APIs for internal services (Express) and front-end tools (React, Highcharts, D3) that the sales team could use on customer calls to drive conversation, suggest insights, and facilitate upsells.

University of Texas at Austin

As a researcher, I divided my time between conducting research in theoretical astrophysics and exoplanet detection, building libraries tooling for the astronomical community, and devising apps and games for astronomy education curricula.

- Led the data analysis effort for the Lick-Carnegie science collaboration (~20 scientists across the United States). Analyzed high-value time series data captured with Keck, APF and Lick telescopes. Created an IDE and libraries for modeling this data, Systemic, which have been used to discover more than 40 new planetary systems by multiple teams. Developed high-performance code packages to simulate planetary formation on UT supercomputer clusters.
- Funded to develop *Super Planet Crash*, an HTML5/JS game that was played more than 15 million times and covered by The Verge, IO9, Huffington Post, and others; and *Systemic Live*, a web app used at Caltech, UF, UT, MIT, SJSU, Yale, Columbia, and Coursera to teach students about data analysis and modeling.

Skills

Machine Learning: Building supervised and unsupervised classification and regression pipelines via state of the art and custom algorithms; devising high-performance statistical and numerical methods; time series analysis and forecasting. Experienced with sklearn, SparkML, H2O.

Data and Software Engineering: Architecting high-volume Spark ETL and machine learning pipelines on AWS EMR; building DS projects from prototypes to production using R, Scala, Python, SQL, JavaScript, Java, and C; building interactive, rich front-end prototypes.

Soft Skills: Mentoring and advising (references available on request); scoping complex tasks, evaluating risk and impact, and acting as a data advocate across multiple teams; working on knotty research problems with tenacity and a positive attitude; explaining complex, multivariate concepts to stakeholders with clarity and empathy.

Education

Doctor of Philosophy (Astronomy & Astrophysics, 2012), University of California at Santa Cruz. Received class year's prize for highest achievement in research, coursework, and teaching.

Master of Science (Astronomy, with highest honors), University of Bologna.

Published 8 first-author refereed publications on time series analysis, numerical optimization, and Monte-Carlo simulations (cited 437 times); a total of 17 refereed publications (cited 1,444 times).

Reference contact information is available on request.