

# **Session summary:**

Machine Learning and Data  
Assimilation in Heliophysics:  
Capturing the Current Picture

## **Three questions of focus for the session:**

- 1) What role are the ML and DA techniques currently playing in the enhancement of Space Weather prediction capabilities?
- 2) What types of broadly-defined problems in Heliophysics (beyond the direct application to Space Weather prediction) require the involvement of the ML and DA techniques?
- 3) What are the current demands in ML-ready data sets and related quality standards in the community?

## **Two scene-setting speakers:**

- Ruizhu Chen (Stanford) “Mapping the Sun’s Far-Side Magnetic Flux from Near-side Helioseismic Measurements by Deep Learning”
- Alberto Sainz Dalda (BAERI/LMSAL) “Machine Learning and Data Assimilation in Heliophysics: Capturing the Current Picture”

# Discussion topics (part 1)

## Quality of the training data

- What limits do the uncertainties in the training data imply on your ML model?

## Transferability of learning

- How to apply the algorithm trained on one type of the data set to another data?

## Interpretability of models

- Can we understand how the ML makes the decisions, and what can we learn from the ML models?

# Discussion topics (part 2)

## ML-ready data preparation

- Is there a demand in the community for the data sets ready for ML applications?

## Ability of ML models to generalize

- How to ensure that the model captures the patterns in the data but not your training data set?

## Assessment of complex models

- How to ensure that the ML solution is “close” to the ground truth, and how to define the closeness?

# Conclusion

We do not necessarily need ML to do great science

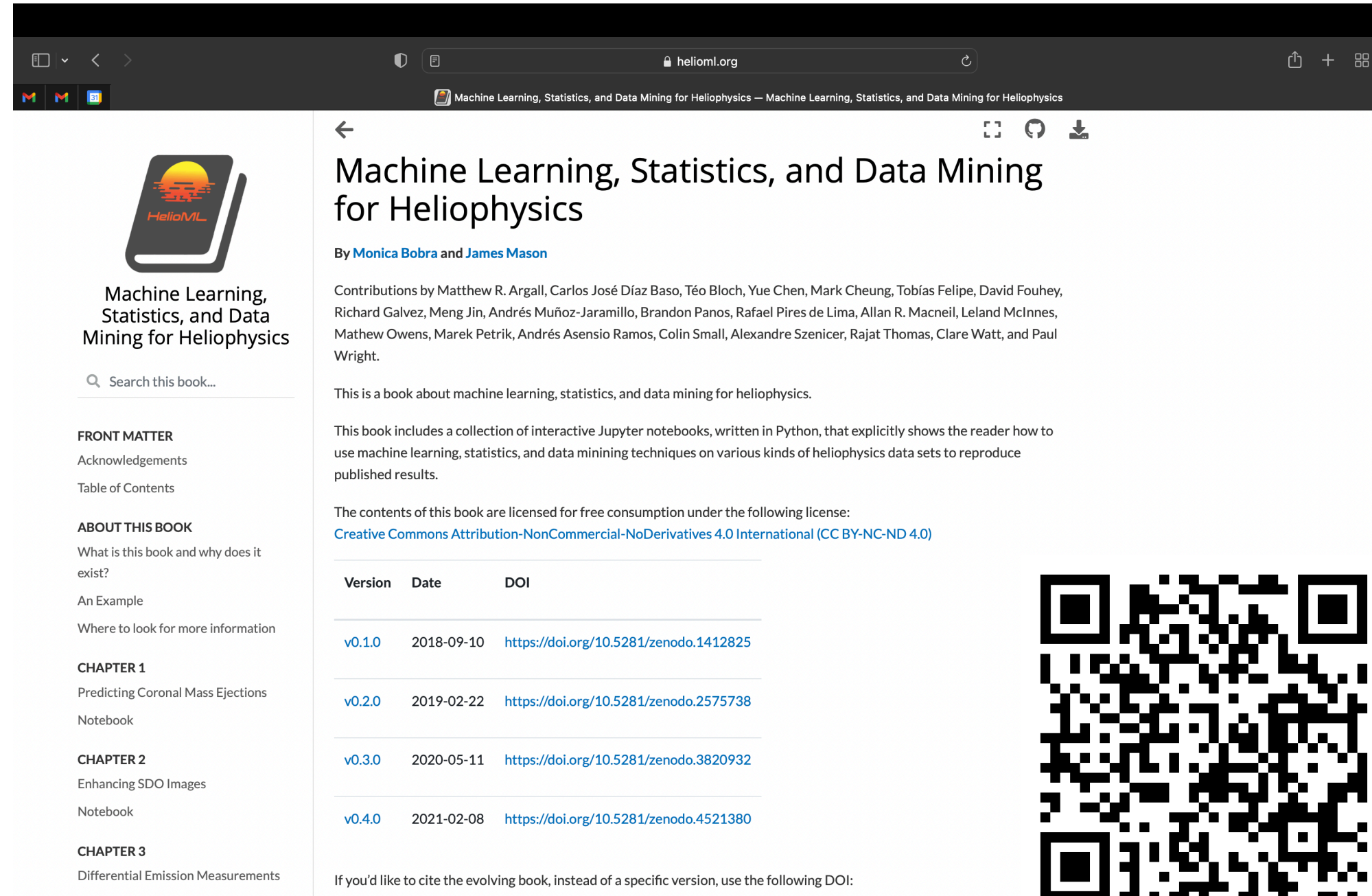
However, ML is a great (and, in some cases, necessary) tool to do a great science

Please find the resources kindly shared by Alberto Sainz Dalda in the following slides

# MACHINE LEARNING AND DEEP LEARNING FOR THE SUN



# RESOURCES



**Machine Learning, Statistics, and Data Mining for Heliophysics**

By [Monica Bobra](#) and [James Mason](#)

Contributions by Matthew R. Argall, Carlos José Díaz Baso, Téó Bloch, Yue Chen, Mark Cheung, Tobias Felipe, David Fouhey, Richard Galvez, Meng Jin, Andrés Muñoz-Jaramillo, Brandon Panos, Rafael Pires de Lima, Allan R. Macneil, Leland McInnes, Mathew Owens, Marek Petrik, Andrés Asensio Ramos, Colin Small, Alexandre Szenicer, Rajat Thomas, Clare Watt, and Paul Wright.

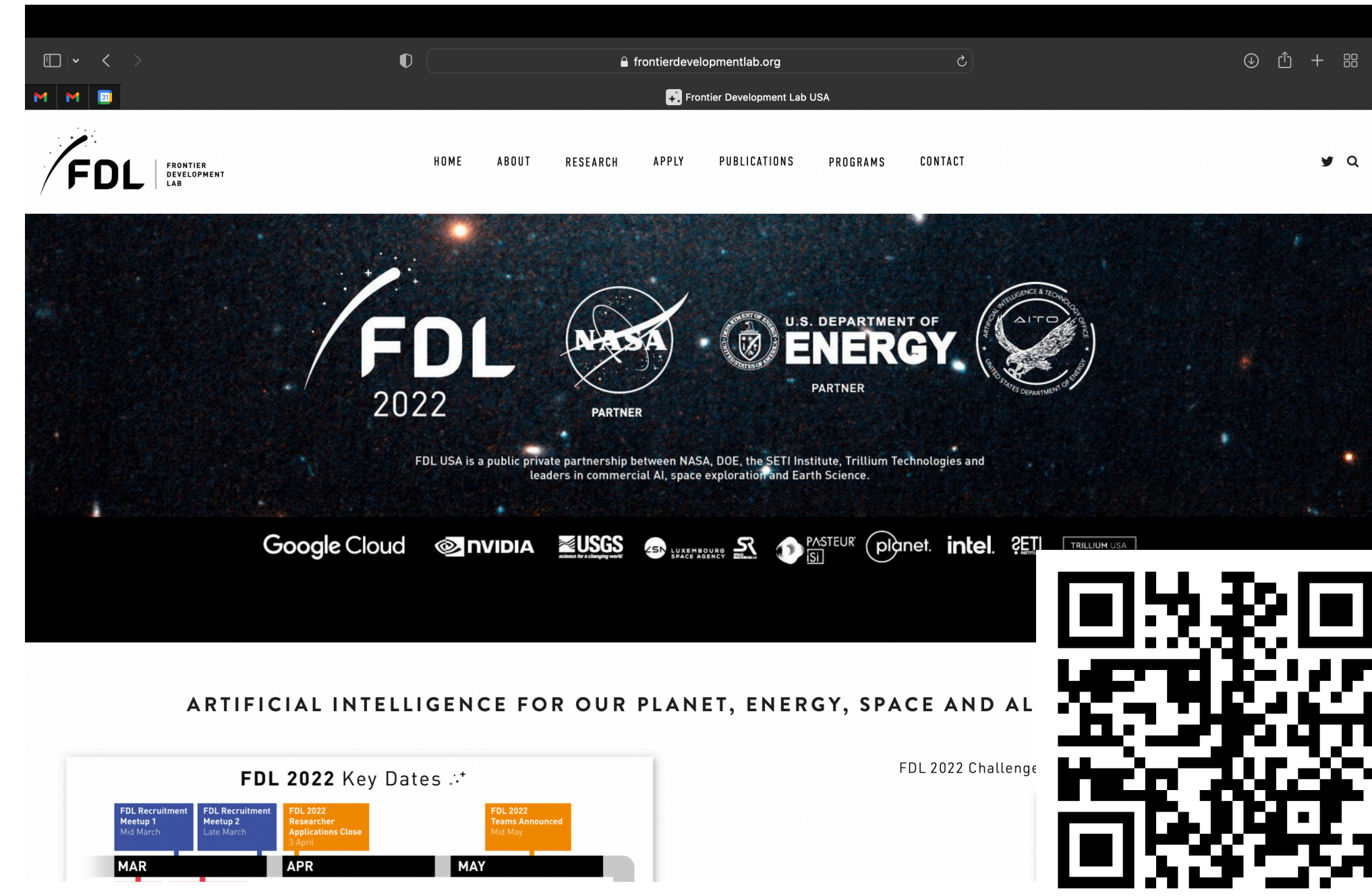
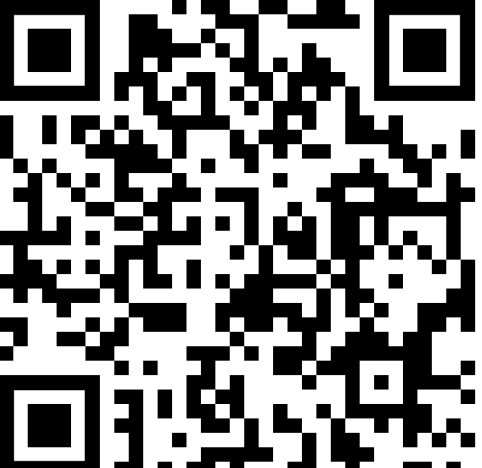
This is a book about machine learning, statistics, and data mining for heliophysics.

This book includes a collection of interactive Jupyter notebooks, written in Python, that explicitly shows the reader how to use machine learning, statistics, and data mining techniques on various kinds of heliophysics data sets to reproduce published results.

The contents of this book are licensed for free consumption under the following license:  
[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](#)

Version	Date	DOI
<a href="#">v0.1.0</a>	2018-09-10	<a href="https://doi.org/10.5281/zenodo.1412825">https://doi.org/10.5281/zenodo.1412825</a>
<a href="#">v0.2.0</a>	2019-02-22	<a href="https://doi.org/10.5281/zenodo.2575738">https://doi.org/10.5281/zenodo.2575738</a>
<a href="#">v0.3.0</a>	2020-05-11	<a href="https://doi.org/10.5281/zenodo.3820932">https://doi.org/10.5281/zenodo.3820932</a>
<a href="#">v0.4.0</a>	2021-02-08	<a href="https://doi.org/10.5281/zenodo.4521380">https://doi.org/10.5281/zenodo.4521380</a>

If you'd like to cite the evolving book, instead of a specific version, use the following DOI:



**FDL 2022**

Partners: NASA, U.S. DEPARTMENT OF ENERGY, SETI Institute, Trillium Technologies, and others.

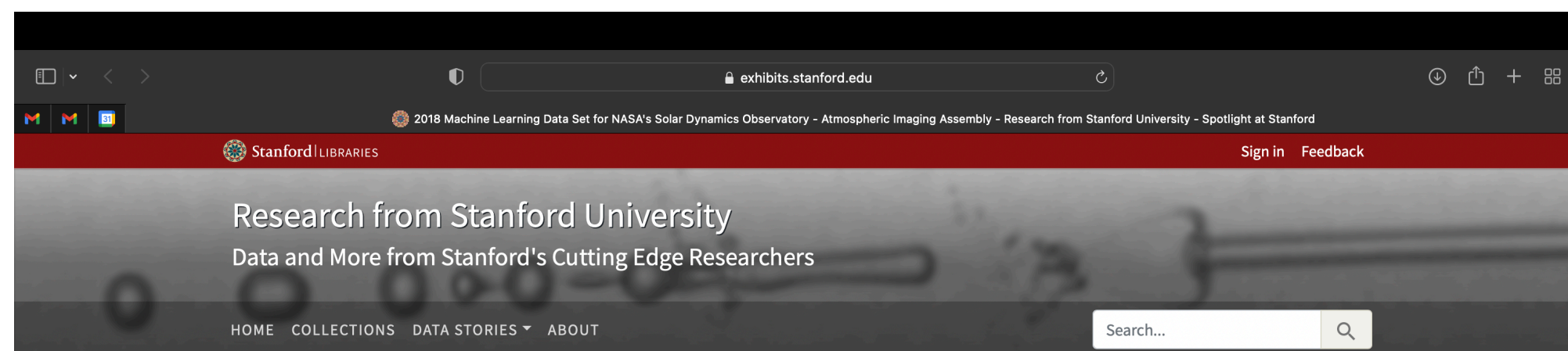

FDL USA is a public-private partnership between NASA, DOE, the SETI Institute, Trillium Technologies and leaders in commercial AI, space exploration and Earth Science.

Artificial Intelligence for our Planet, Energy, Space and All

**FDL 2022 Key Dates**

- FDL Recruitment Meetup 1: Mid-March
- FDL Recruitment Meetup 2: Late-March
- FDL 2022 Researcher Applications Close: 7 April
- FDL 2022 Teams Announced: Mid-May

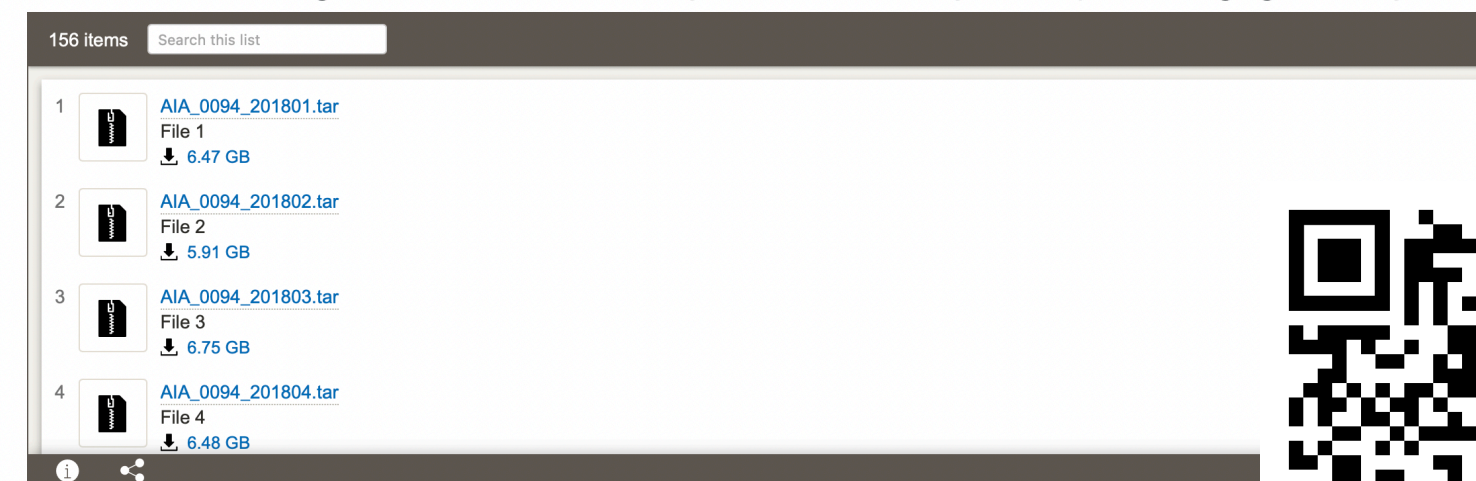
FDL 2022 Challenge



Research from Stanford University  
Data and More from Stanford's Cutting Edge Researchers

HOME COLLECTIONS DATA STORIES ABOUT

## 2018 Machine Learning Data Set for NASA's Solar Dynamics Observatory - Atmospheric Imaging Assembly



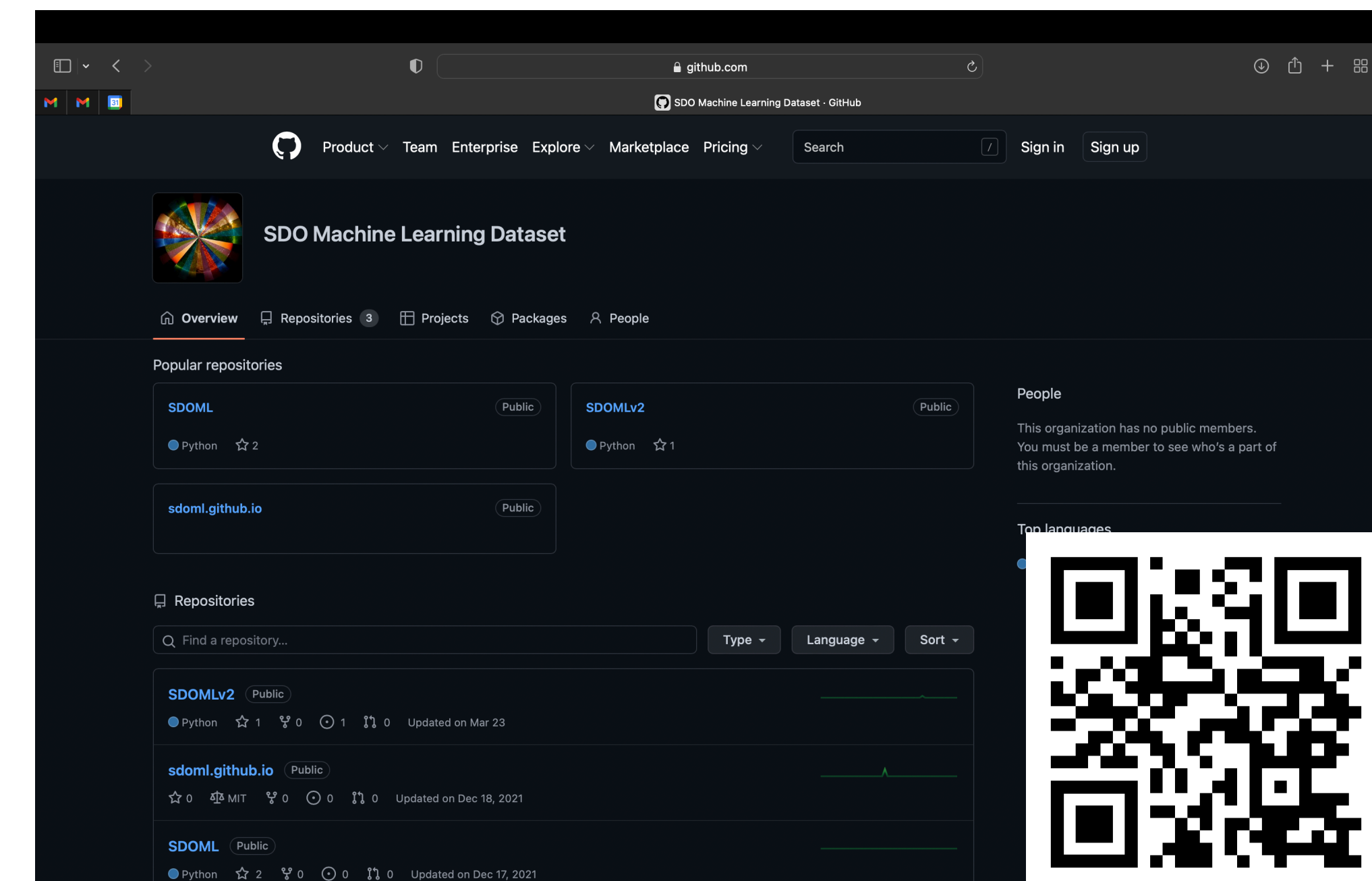
156 items

- AIA\_0094\_201801.tar (File 1, 6.47 GB)
- AIA\_0094\_201802.tar (File 2, 5.91 GB)
- AIA\_0094\_201803.tar (File 3, 6.75 GB)
- AIA\_0094\_201804.tar (File 4, 6.48 GB)

[purl.stanford.edu/nk828sc2920](https://purl.stanford.edu/nk828sc2920)

Author: Fouhey, David, Jin, Meng, Cheung, Mark, Munoz-Jaramillo, Abndres, Galvez, Richard, Thomas, Rajat, Wright, Paul, Szenicer, Alexander, Bobra, Monica G., Liu, Yang, and Mason, James

Description: We present a curated dataset from the NASA Solar Dynamics Observatory (SDO) mission in a format suitable for machine learning research. Beginning from level 1 scientific products we have processed unique instrumental corrections, downsampled to manageable spatial and temporal resolutions, and



SDO Machine Learning Dataset

Overview Repositories (3) Projects Packages People

Popular repositories

- SDOML (Public, Python, 2 stars)
- SDOMLv2 (Public, Python, 1 star)
- sdoml.github.io (Public)


Repositories

- SDOMLv2 (Public, Python, 1 star, 0 forks, 0 issues, 0 pull requests, updated on Mar 23)
- sdoml.github.io (Public, Python, 0 stars, 0 forks, 0 issues, 0 pull requests, updated on Dec 18, 2021)
- SDOML (Public, Python, 2 stars, 0 forks, 0 issues, 0 pull requests, updated on Dec 17, 2021)

People

This organization has no public members. You must be a member to see who's a part of this organization.

Top Languages



# RESOURCES

XXX Canary Islands Winter School of Astrophysics  
La Laguna, Tenerife, Spain - 4-10 November 2018

## Big Data analysis in astronomy

### Presentation

The XXX Canary Islands Winter School of Astrophysics, organized by the Instituto de Astrofísica de Canarias (IAC), focusses on **Big Data Analysis in Astronomy**. The School, to be held in San Cristobal de La Laguna (Tenerife, Spain) from 4 to 10 November 2018, will admit around 60-70 PhD students and young Post-Docs. In a relaxed environment, the School will give participants the opportunity to learn from world-renowned specialists about the techniques used in Astronomy in the analysis of Big Data. Participants of the Winter School will have the opportunity to highlight their current work by presenting a poster. The duration of the school, including the visits to the observatories, is equivalent to a total of 30 hours (3.0 academic credits). Deadline for both the Registration and Applications is **17th June 2018**.

### Scientific Rationale

We are witnessing an unprecedented growth in the volume and complexity of astronomical data instruments are providing massive amount of raw data, images and spectra. This situation will be future, in particular with the upcoming telescopes LSST, SKA, and Euclid. In the era of 'Big Data', will need to deal routinely with tera- and even petabytes of information. The analysis of these data in the traditional way, and key decisions in the process will have to rely on numerical algorithms. digesting data and extracting the relevant information from them will pose significant challenges technical and at the analytical level. Machine learning techniques, which belong to the realms of mathematics and statistics, provide key tools to be used for the task. Acknowledging the important revolution from an early stage, the XXX Canary Islands Winter School will allow the students to be current developments in this field, and with the new techniques to be used. The Winter School will include lectures and practical work focussed on dealing with big astronomical data sets. A particular focus will be on learning techniques.

### Organizing Committee

Antonio M. Pérez (Co-Director of the School)

GitHub - dalya/IAC\_Winter\_School\_2018: The lectures and the tutorials given in IAC Winter School 2018 on unsupervised learning

master 1 branch 0 tags

Go to file Code

Update README.md 4 years ago

data/stellar\_spectra\_sample Delete temp\_file.txt 4 years ago

presentations Create temp\_file.txt 4 years ago

README.md Update README.md 4 years ago

clustering\_algorithms.ipynb Add files via upload 4 years ago

dim\_reduction\_pca\_tsne.ipynb Add files via upload 4 years ago

supervised\_learning\_and\_outlier\_detection.ipynb Add files via upload 4 years ago

## README.md

### IAC Winter School 2018: unsupervised machine learning algorithms in Astronomy

The XXX Winter School of Astrophysics on Big Data in Astronomy: <http://www.iac.es/winterschool/2018/>.

This repository contains the slides from the lectures I gave during the school, and python tutorials. The topics I covered in the school:

- Introduction to Unsupervised Learning:** The main differences between supervised and unsupervised learning algorithms, and the basic anatomy of unsupervised learning algorithms.

irisreader documentation - irisreader 0.3 documentation

Docs » irisreader documentation View page source

## irisreader documentation

The IRISreader library provides functionality to read level 2 data from the **IRIS satellite** and to preprocess it for big data applications.

The following structure contains a quickstart tutorial and a reference library of the irisreader interface. IRISreader also involves a lot of functionality that is hidden from the end user, these classes and methods have not been documented yet and are not intended for use by non-developers.

### Contents:

- Installation
- Tutorial
- Configuration
  - irisreader.config\_template
- Reading
  - irisreader.observation
  - irisreader.obs\_iterator
  - irisreader.get\_lines
  - irisreader.has\_line
  - irisreader.sji\_cube
  - irisreader.raster\_cube
  - irisreader.get\_obs\_path
- Preprocessing
  - irisreader.preprocessing.image\_cropper
  - irisreader.preprocessing.image\_cube\_cropper
  - irisreader.preprocessing.spectrum\_interpolator
- Co-alignment
  - irisreader.coalignment.find\_closest\_sji
  - irisreader.coalignment.find\_closest\_raster
  - irisreader.coalignment.goes\_data
  - irisreader.coalignment.hek\_data
- Utils

Share your knowledge...  
(type here)