



**STScI** | SPACE TELESCOPE  
SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

# Enabling Data Science at MAST

---

Josh Peek for MAST

**STScI Town Hall**

**AAS 240, Pasadena**



*Data science, including applications of machine learning, will play an increasing role in astronomical research over the coming decade. Incorporating training in this area at the graduate level and beyond will better prepare researchers regardless of whether they pursue careers in astrophysics or in other STEM fields.*





# MAST as a library: a place for data, tools, teaching, & gathering





TIME SERIES

INTEGRATED

KNOWLEDGE

ENGINE



TIME SERIES

INTEGRATED

KNOWLEDGE

ENGINE



## The TIKE is live now!

---

<https://timeseries.science.stsci.edu>

A JupyterHub platform for exploring TESS, Kepler & K2 timeseries data

Provides tools for accessing MAST AWS Open Data

Full of notebooks and software for this kind of exploration



# The TIKE is a full, free-to-use software environment for everyone

The screenshot displays the TIKE interface with three main components:

- File Browser (Left):** Shows a sidebar with a search bar and a file list. The list includes folders for 'Desktop' (modified 21 minutes ago), 'spaceteles...' (3 months ago), and 'tike\_content' (3 months ago).
- Jupyter Notebook (Middle):** The active notebook is 'data-access.ipynb'. It contains a code cell with a function call: `pixelfile = search[0].download(download_dir=".")`. Below the code, the text explains that the `enable_cloud_dataset()` function is used to download data from AWS instead of MAST. A section titled '3. Using the AWS command-line tool' describes how to use the `aws s3` CLI to access data in S3. A code cell shows the command: `!aws s3 ls s3://stpubdata/tess/public/ --no-sign-requirement`. The text further explains the directory structure and provides an example location for TESS Sector 11 data.
- Terminal (Right):** A terminal window titled 'Terminal 1' shows the output of the `pip install healpy` command. The output lists various requirements that are already satisfied in the environment, such as `matplotlib`, `numpy`, `scipy`, `astropy`, `PyYAML`, `pyerfa`, `packaging`, `python-dateutil`, `kiwisolver`, `fonttools`, `cycler`, `pyarsing`, `pillow`, and `six`. The final output is: `Installing collected packages: healpy`, `Successfully installed healpy-1.15.2`, and `(tess) jovyan@notebook:~$ pip install healpy`.

NUMPY, SCIPY, MATPLOTLIB, PANDAS...

ASTROPY, ASTROQUERY, PYVO...

LIGHTKURVE, ASTROCAT, EVEREST...

EMCEE, GEORGE, CELERITE...

TENSORFLOW, SCIKIT-LEARN...

AWSCLI, BOTO3, S3FS...





# The TIKE is loaded with dozens of science examples for TESS & Kepler

The screenshot displays the TIKE JupyterLab interface. On the left is a file browser with a search bar and a table of files. The main area shows a notebook with the title 'science-examples.md' containing text and a list of links.

Name	Last Modified
Desktop	3 minutes ago
spacetelescope-notebooks	3 months ago
tike_content	3 months ago

## Example science tutorials

These tutorials are specific to a particular science case and show how to use the MAST programmatic interface and the JupyterHub science platform to do research. The links on this page may only work within the TIKE platform, but you can find external links to the notebooks in GitHub here: [Example science tutorials \(GitHub\)](#).

**NOTE!** In order for these notebooks to access the expected packages in TIKE, you may have to switch to the TESS kernel after you open them. You can do this either by using the dropdown menu at the upper right of the notebook, or in the top JupyterLab menu (Kernel > Change Kernel...).

## Science with TESS

- [Beginner: Read and Plot A TESS Data Validation Timeseries File](#)
- [Beginner: Read and Display a TESS Full Frame Image](#)
- [Beginner: Read and Plot A TESS Light Curve File](#)
- [Beginner: Read and Display A TESS Target Pixel File](#)
- [Beginner: Search The TESS Input Catalog Centered On HD 209458](#)
- [Beginner: A Tour of the Contents of the TESS 2-minute Cadence Data](#)
- [Beginner: Cutout of the TESS FFIs using Astrocut and Astroquery](#)
- [Intermediate: Search and Download GI Program Light Curves](#)
- [Intermediate: Create TESS FFI Cutout using Python Requests](#)

## Science with Kepler & K2

### Beginner Notebooks

- [Using Kepler Data to Plot a Light Curve](#)





## The TIKE accelerates Big Data analyses with the cloud

*Time to retrieve 30 TESS images (FFIs)*

Where you are	Where you get the data	Time
Home Wifi	Baltimore	14 minutes
Home Wifi	Cloud	6 minutes
TIKE	Baltimore	1 minute
TIKE	Cloud	<b>9 seconds</b>





# Learn More about TIKE from the creators



## The Timeseries Integrated Knowledge Engine (TIKE): Millions of Kepler/TESS exoplanet light curves, Zero Downloads, Zero Installations

Gregory Snyder<sup>1</sup>, Geert Barentsen<sup>2</sup>, Scott Fleming<sup>1</sup>, Susan Mullally<sup>1</sup>, Jason Tumlinson<sup>1</sup>, MAST Team<sup>1</sup>, Science Platforms Team<sup>1</sup>

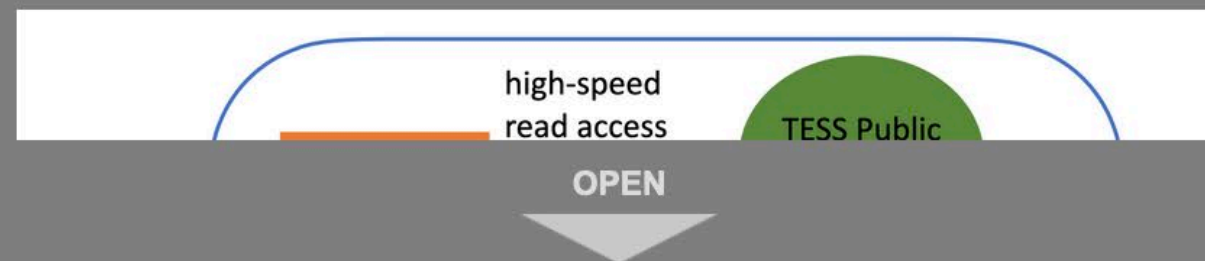
<sup>1</sup>STScI, <sup>2</sup>NASA Ames

See Greg's Poster!

### Access MAST Data in AWS

[bit.ly/tike-data](https://bit.ly/tike-data)

MAST AWS Open Data are free to access with no AWS account needed!



### TIKE JupyterHub Platform

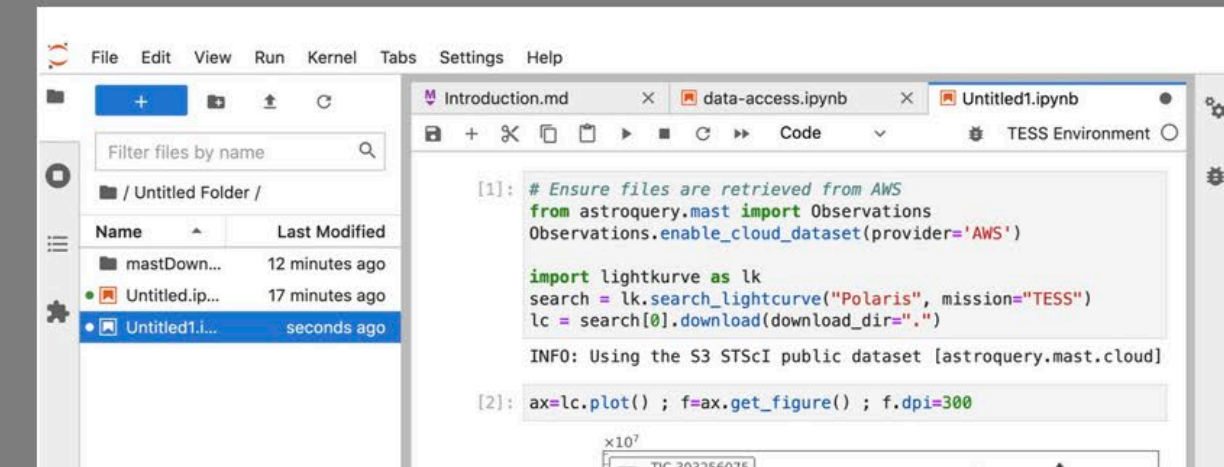
Available Now:

<https://timeseries.science.stsci.edu>

- Log in with your MyST Account
- Run 20+ pre-installed software packages
- Compute near [MAST AWS Open Data](#)
- Try Tutorials and example notebooks



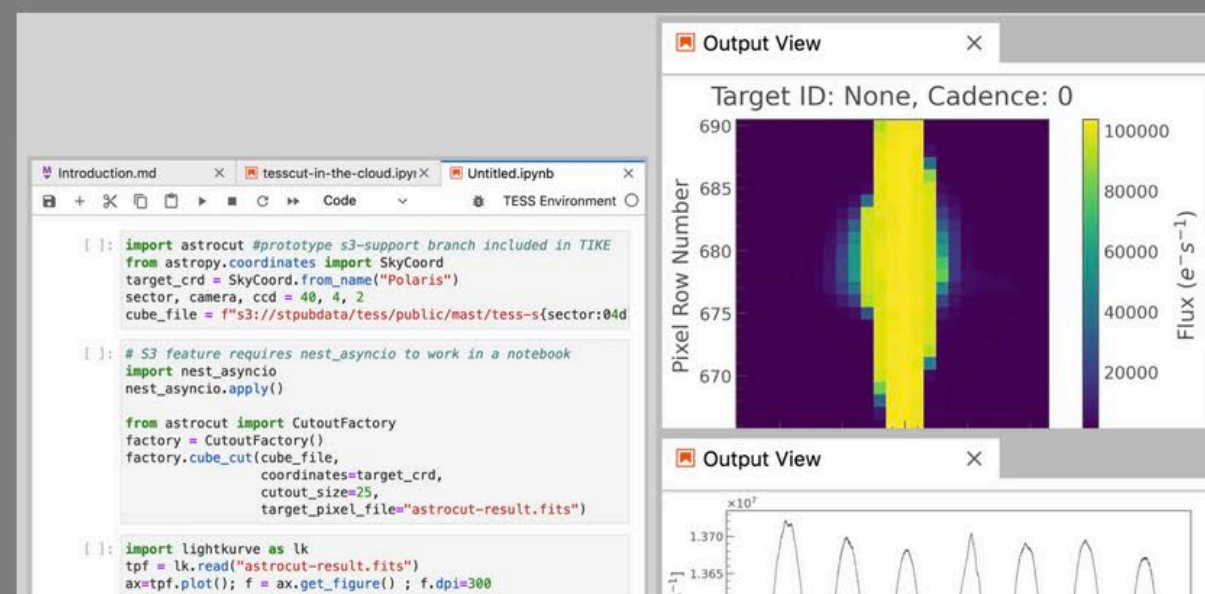
### Quickly Visualize and Analyze Light Curves



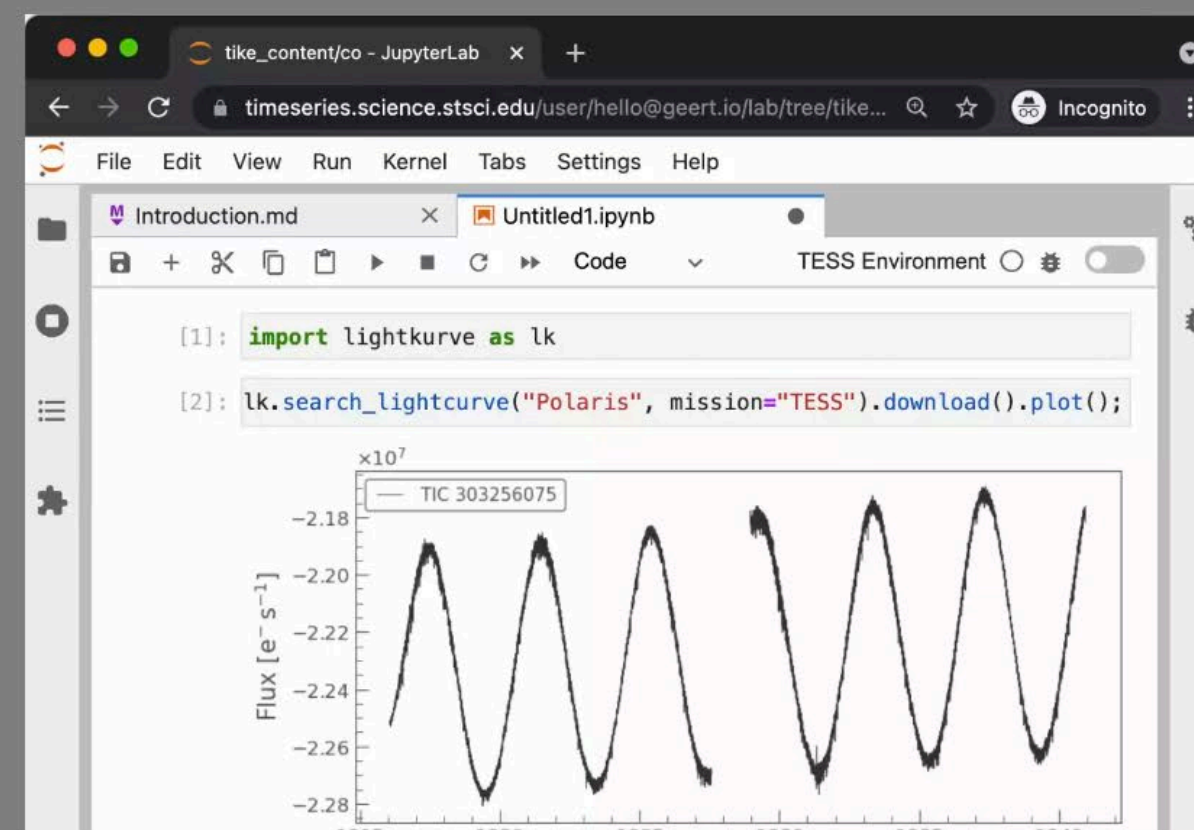
OPEN

### Cut Out TESS FFI Cubes in the Cloud

[bit.ly/tike-cutout](https://bit.ly/tike-cutout)



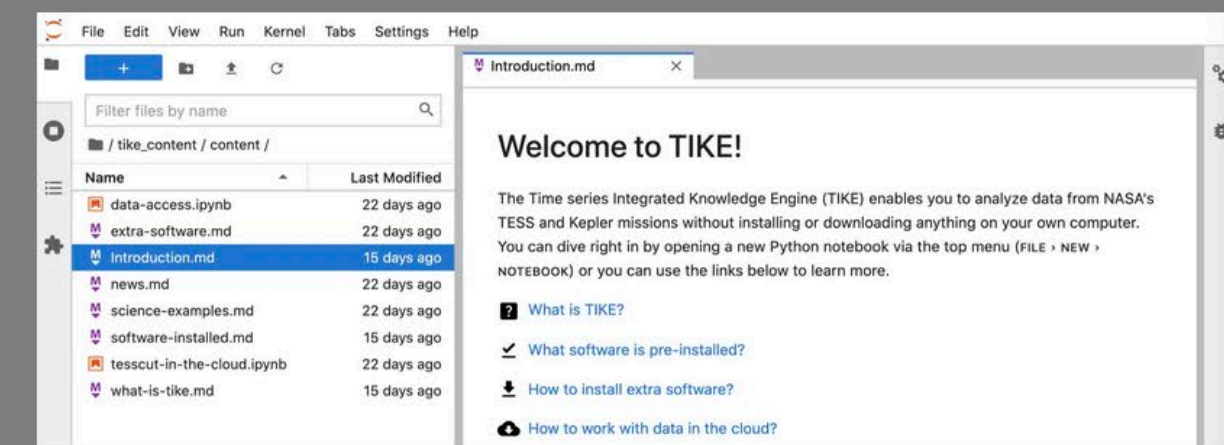
OPEN



OPEN

### Get Help and Give Feedback

Email [archive@stsci.edu](mailto:archive@stsci.edu) or file an archive helpdesk ticket at <https://stsci.service-now.com/mast>



OPEN

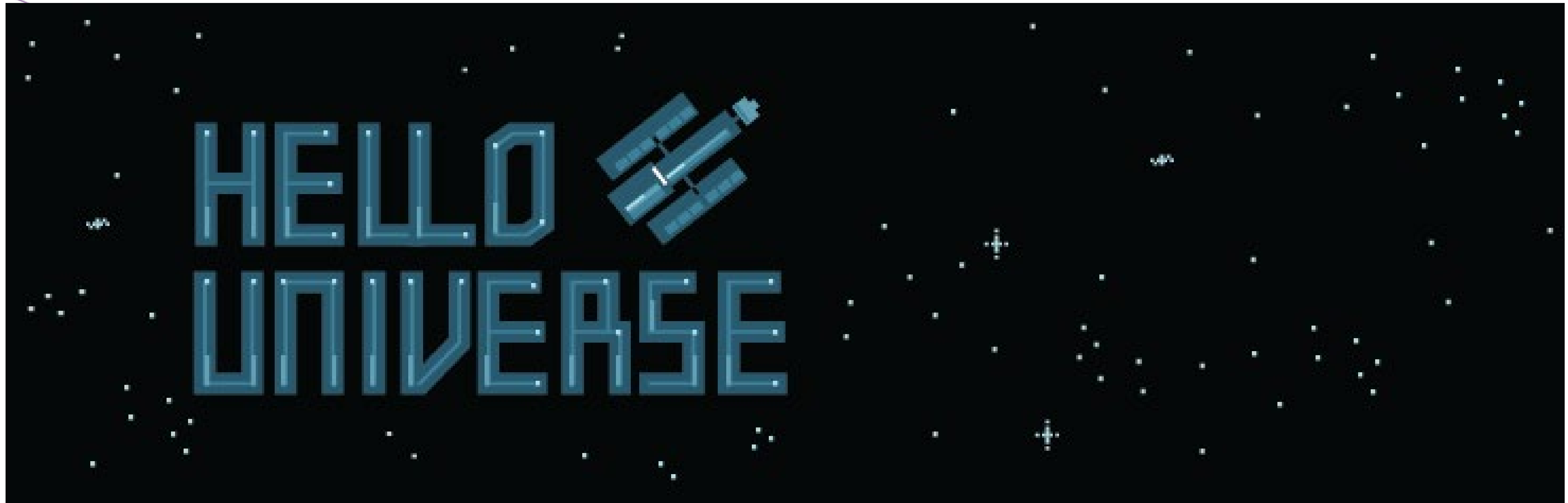


HELLO   
UNIVERSE





## We developed *Hello Universe* to make ML more accessible in astro



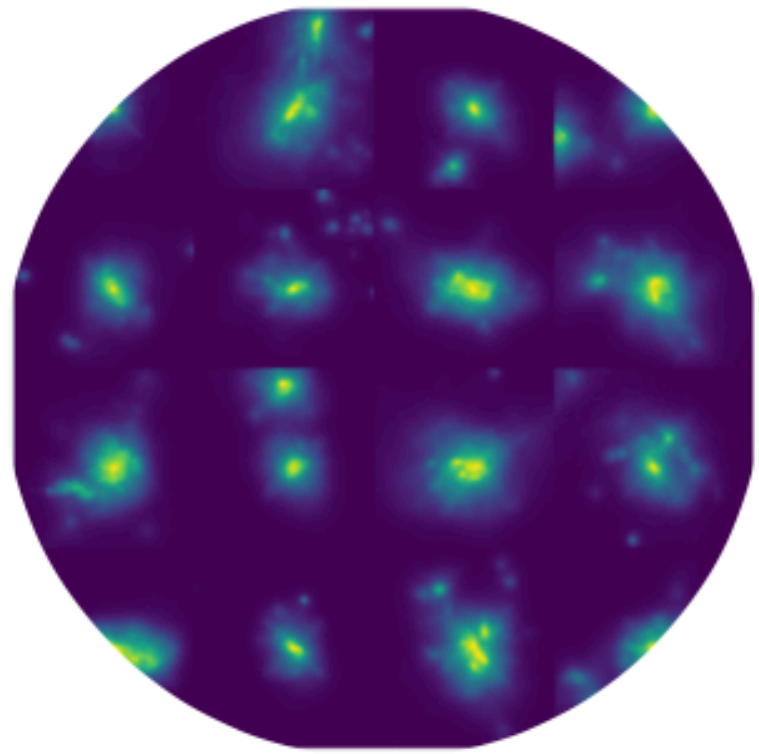
Upcoming large and rich data sets will require astronomers to develop fluency in data science methods. Hello Universe addresses this need by providing:

- Data: reliable **high level science products** (HLSPs) that are appropriate for testing and benchmarking new machine learning algorithms.
- Code: pedagogical, **always working** Jupyter notebooks that provide a step-by-step example of how to apply machine learning (ML) to the data.





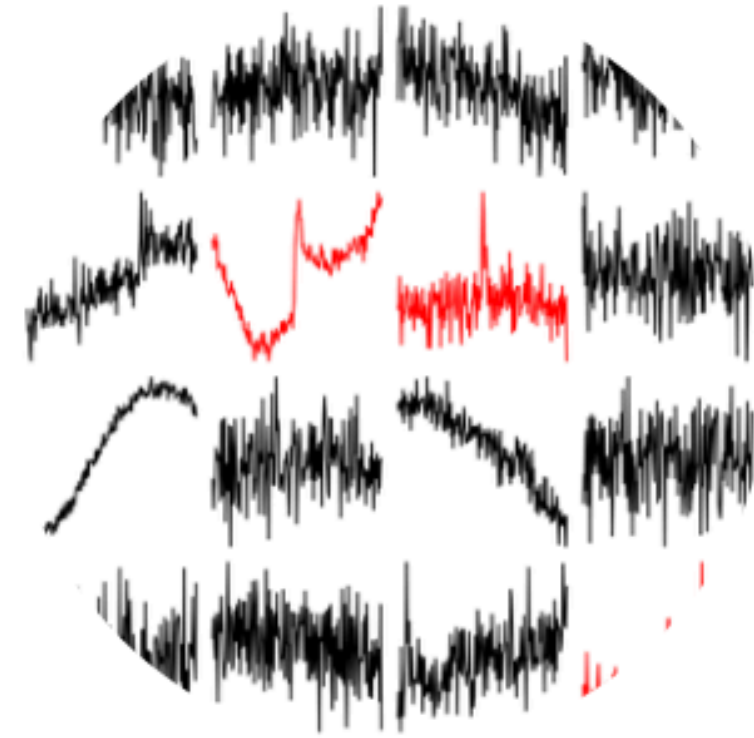
# Hello Universe launching today with 4 notebooks + data sets



## Classifying JWST/HST galaxy mergers with CNNs

neural networks | 2d data | classification | overfitting | confusion matrix

[Read More](#)



## Classifying TESS stellar flares with CNNs

neural networks | 1d data | classification | prediction

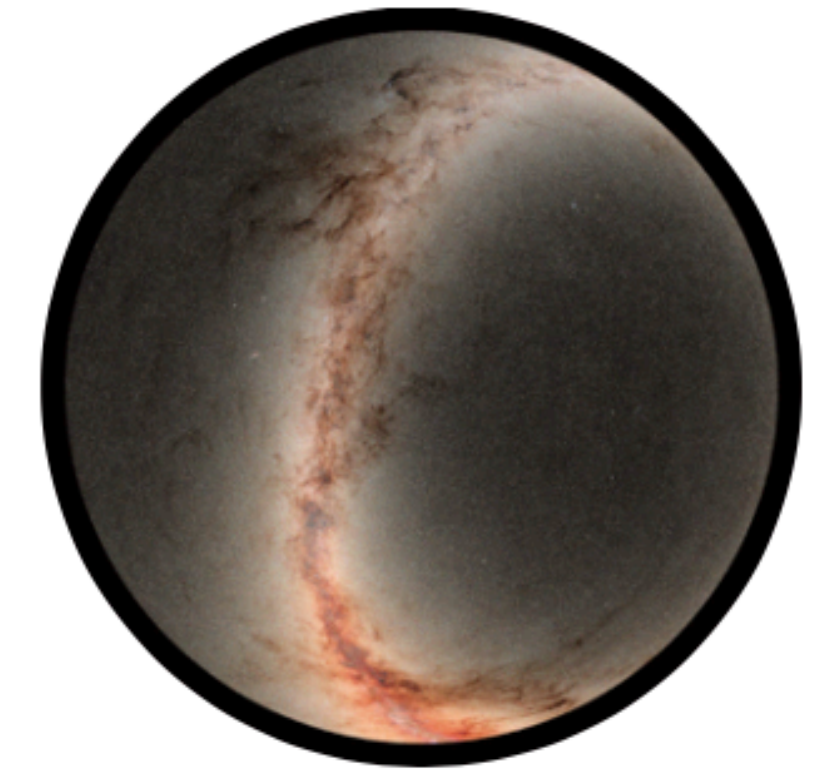
[Read More](#)



## Predicting 3D-HST redshift with decision trees

decision trees | 1d data | regression | cross-validation

[Read More](#)



## Classifying Pan-STARRS with (un)supervised learning

classification | 1d data | PCA | tSNE | k-means | SGD | unsupervised | supervised

[Read More](#)





# Hello Universe launching today with 4 notebooks + data sets

## Classifying galaxy mergers with JWST/HST and CNNs

This entry explores how to classify merging galaxies vs. non-merging galaxies from multi-wavelength imaging from the James Webb Space Telescope (JWST) and the Hubble Space Telescope (HST) with convolutional neural networks (CNNs). This approach was used by the [DeepMerge](#) team ([Ciprijanovic et al. 2020](#)) to classify mergers in synthetic observations of simulated galaxies. The same workflow is presented in simplified format to walk through the construction of the CNN model. The results are then validated and the performance is discussed.

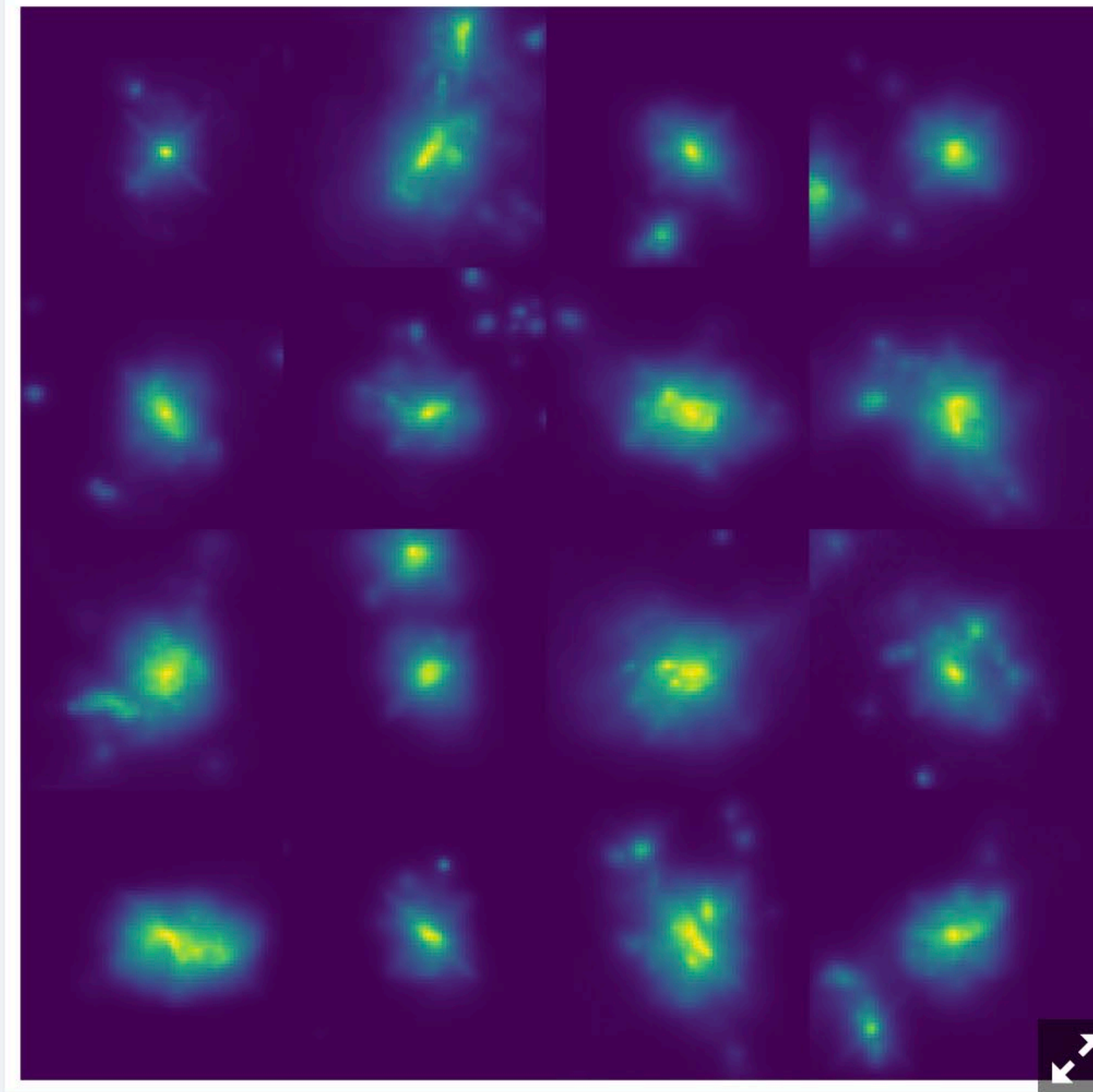
**Data:** [The DeepMerge HLSP](#)

**Notebook:** [Classifying JWST-HST galaxy mergers with CNNs](#)

**Released:** 2022-06-12

**Updated:** 2022-06-12

**Tags:** classification, deep learning, 1d data



Example simulated galaxy images (JWST NIRCам F356W) from [DeepMerge](#) ([Ciprijanovic et al. 2020](#)).





# Hello Universe is already helping astronomers new to ML

See  
Jonah's  
Poster!



## Star Clusters on FIRE: Classifying Star Clusters in Synthetic Galaxy Images using Machine Learning

Jonah Otto<sup>1, 2</sup>, Sarah Loebman<sup>2</sup>, Matt Orr<sup>3</sup>, Josh Peek<sup>4</sup>, Andrew Wetzel<sup>5</sup>, Binod Bhattacharai<sup>2</sup>, Dylan Benton<sup>2</sup>, Micah Oeur<sup>2</sup>, Erick Villegas<sup>2</sup>

<sup>1</sup>California State University, Los Angeles <sup>2</sup>University of California, Merced <sup>3</sup>Rutgers University and the Center for Computational Astrophysics <sup>4</sup>Space Telescope Science Institute and Johns Hopkins University <sup>5</sup>University of California, Davis

### Producing Synthetic Galaxy Images



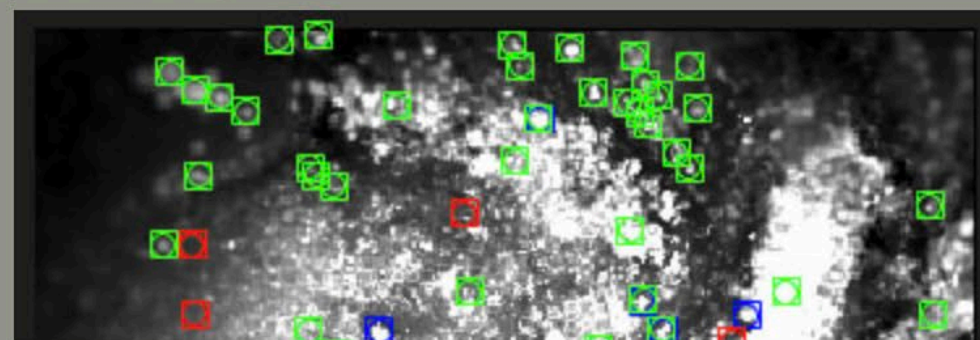
Below is an example synthetic galaxy image, we focused on galaxy m12i of the *Latte* suite in this study.

OPEN

### Source Extractor finds Potential Clusters



Below is an example of potential star clusters identified using the Friends of Friends algorithm, Source Extractor (Bertin & Arnouts 1996) and visually.

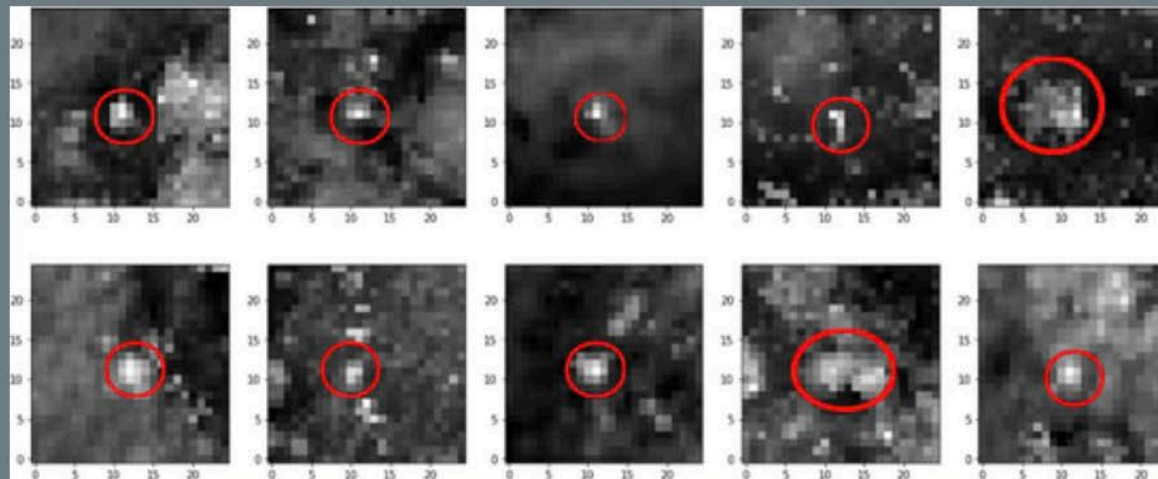


OPEN

### Synthetic Galaxy Star Cluster Catalog



We use machine learning to successfully classify over 35,000 potential star clusters identified in synthetic galaxy images, **thousands of times faster** than humans could at a **similar level of accuracy**.



Above are a random sample of class 1 star clusters identified in m12i

Class 1	Class 2	Class 3	Class 4
2658	10967	5401	16322
1961	4488	9019	19880
3279	4898	6703	20468
1986	7113	8209	18040
1754	6501	12453	14640

The above table shows how our CNN classified the ~35,000 potential star clusters generated from all 555 synthetic images.

OPEN

### Human Generated Training Set Labels



A CNN trained on human generated labels can only be as accurate as the labels it is given.

Following Grasha et al. 2015, each potential star cluster identified by Source Extractor could be labeled as 1 of 4 classes.

OPEN

### Convolutional Neural Network



We developed our own CNN because StarcNet (Perez et al. 2021), a publicly available CNN architecture, was designed with observational data in mind.

#### StarcNet vs Manual Classification

Manual StarcNet



OPEN