

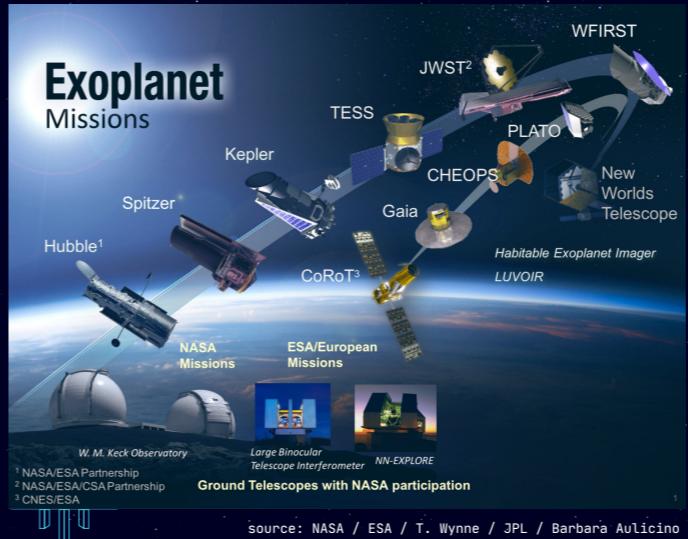




BACKGROUND

EXOPLANETS

- Kepler
- K2
- TESS



Every telescope NASA launches has a certain mission objective: Kepler, K2, and TESS are looking for planets orbiting stars outside our solar system...

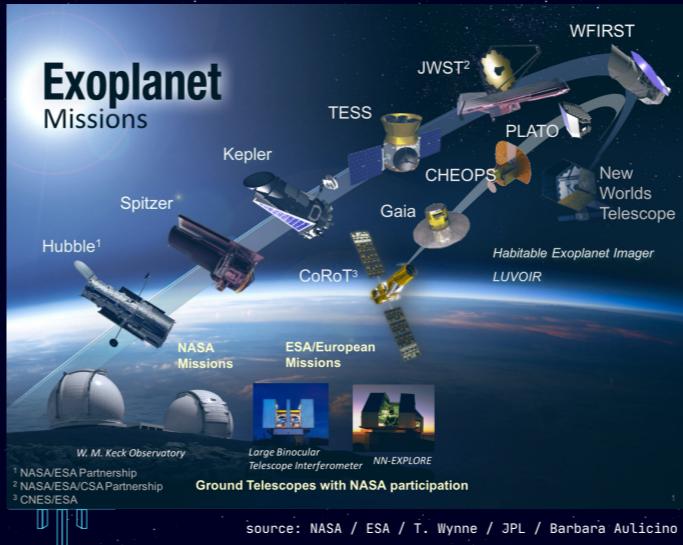
BACKGROUND

EXOPLANETS

- Kepler
- K2
- TESS

PULSARS (DEAD STARS)

- HTRU2



the High Time Resolution Survey allows us to find pulsars (dead stars)...

BACKGROUND

EXOPLANETS

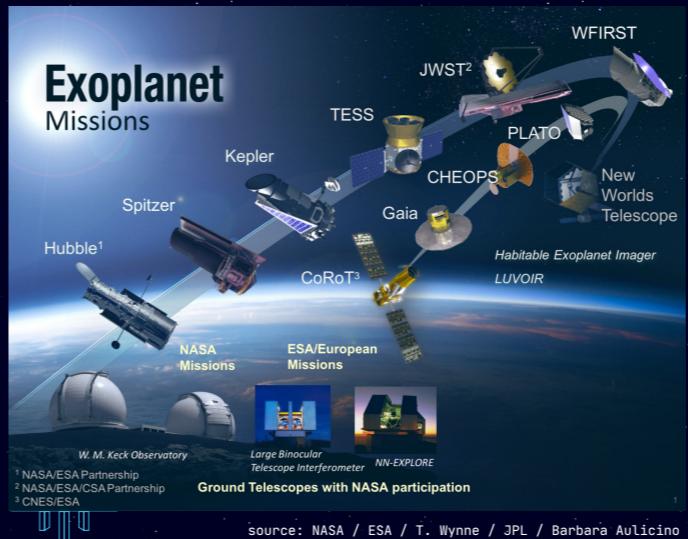
- Kepler
- K2
- TESS

PULSARS (DEAD STARS)

- HTRU2

GALAXY/STAR FORMATION

- JWST



and the James Webb Space Telescope when it launches will also look for early formation of stars and galaxies.

BACKGROUND

EXOPLANETS

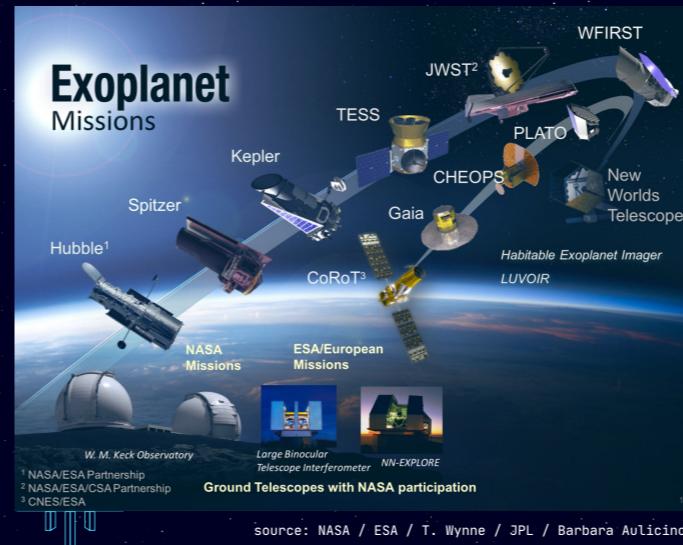
- Kepler: signals - longer time periods
- K2: signals - shorter time periods
- TESS: signals and full frame images

PULSARS (DEAD STARS)

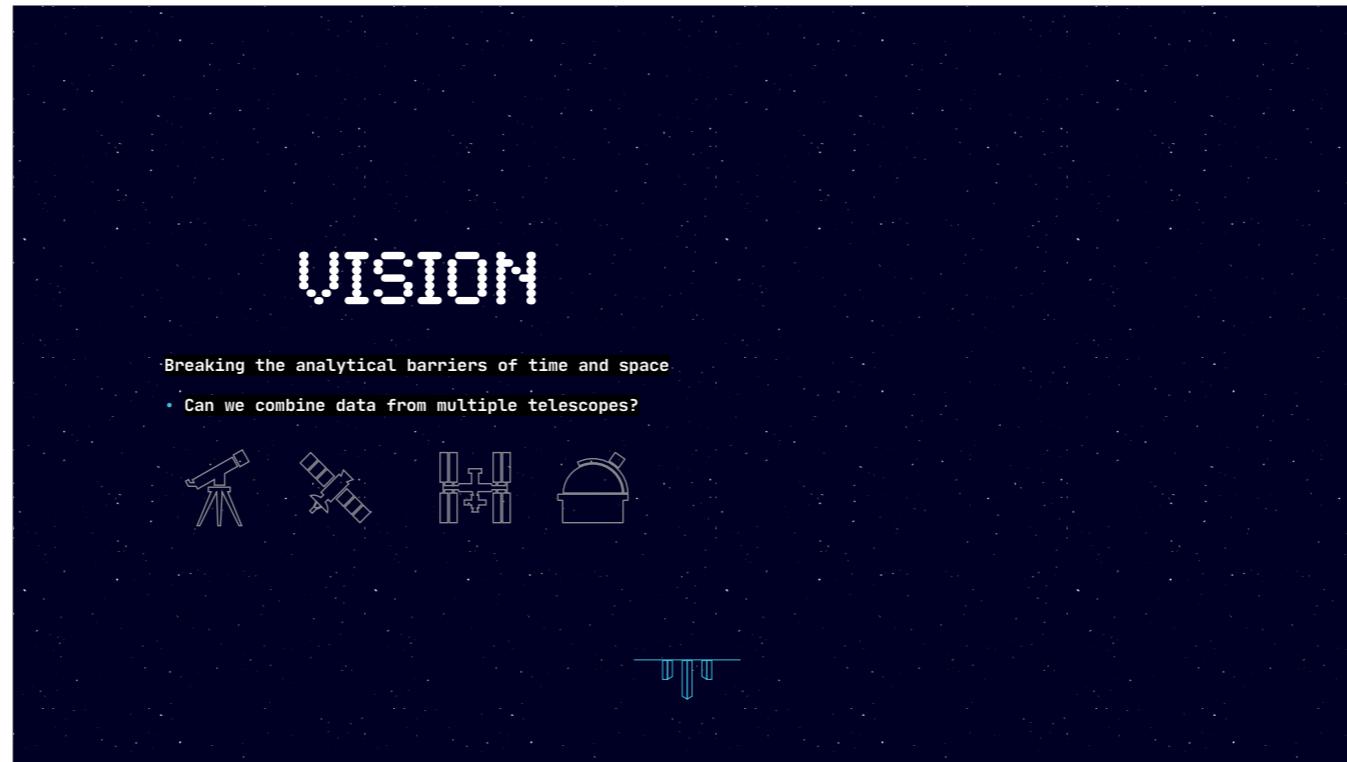
- HTRU2: statistical properties of signals

GALAXY/STAR FORMATION

- JWST: full-frame images



Each telescope is built with specific technology for data capture that allows it to achieve its objective. In hunting for exoplanets using Kepler and K2 data, we can perform signal processing on light flux values from stars over long and short periods of time. With TESS, we can analyze full-frame images.

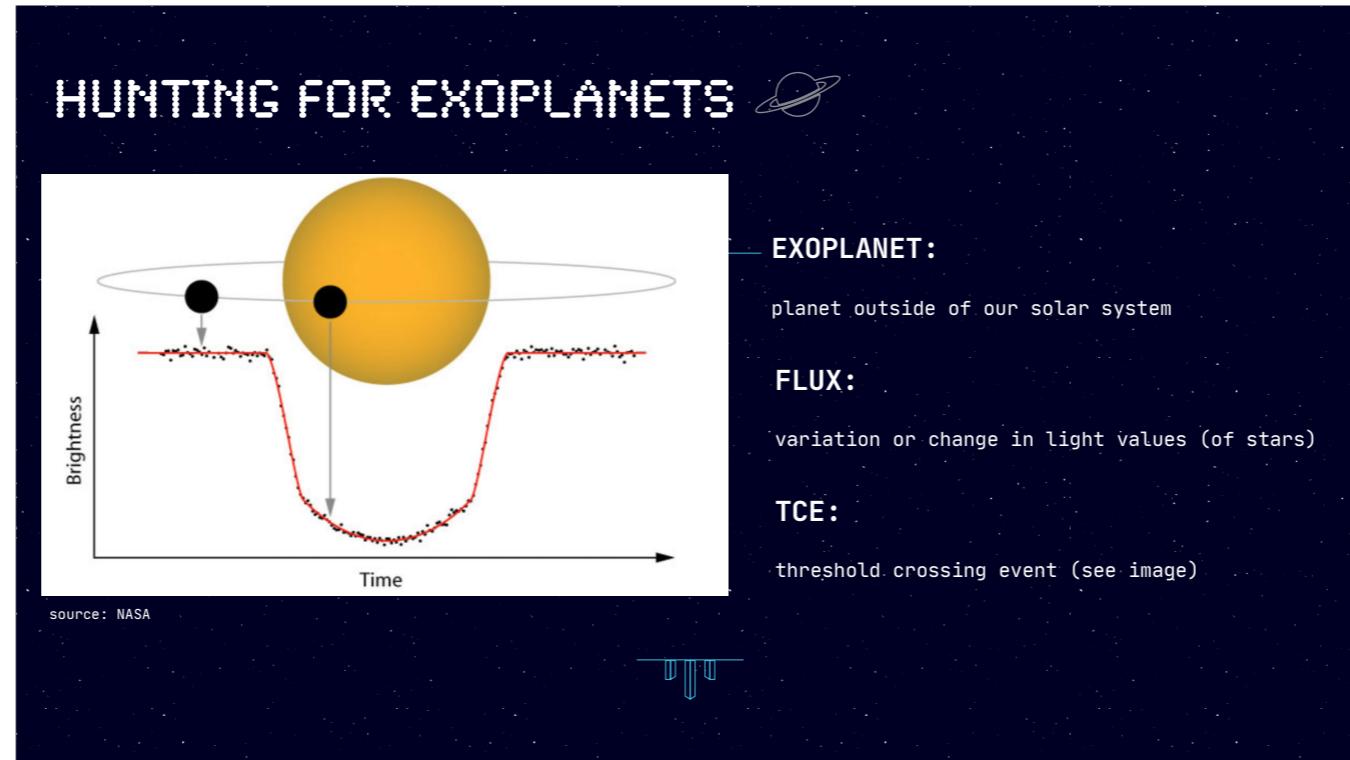


The main vision for STARSKØPE is to build a model that is not limited to analyzing data from just one telescope, but multiple telescopes. This would give us a wider window of time as well as a higher dimension of space (looking at one thing from multiple angles) in which to classify objects. In other words, we can use machine learning to break the barriers of time and space that limit astrophysical object classification to just one telescope, or one campaign. Telescopic becomes cyberoptic, and human intelligence is extended by artificial intelligence.



The mission for starskøpe is to build a Cyberoptic Artificial Telescope for Astrophysical Object Classification.

1. In order to perform USEFUL and ACCURATE machine learning analysis on astrophysical data, we need to avoid “black box” algorithms which prevent us from understanding why the model classified objects the way it did. If we don’t ask the “WHY?” then we aren’t very good scientists.
2. The model also needs to Account for the physics, not just the math. In the most simple terms, if we don’t know and account for units and laws of physics in the equations, we lose enough points to fail the test.
3. The model needs to account for unique attributes and limitations associated with the telescope from which the data was collected,
4. ...as well as correct for any issues and errors that occurred during the campaign that data was being collected.

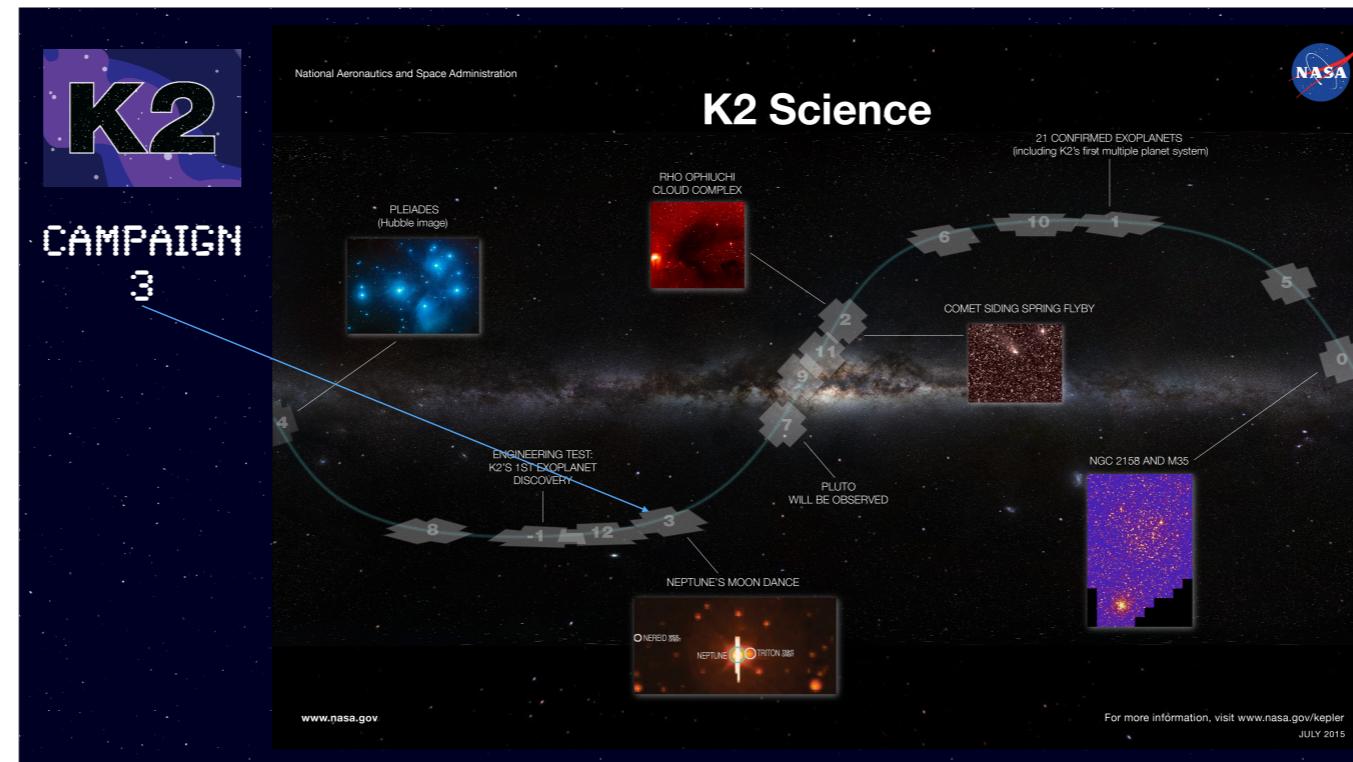


Before taking off, let's define a Couple of terms that may not be familiar to everyone.

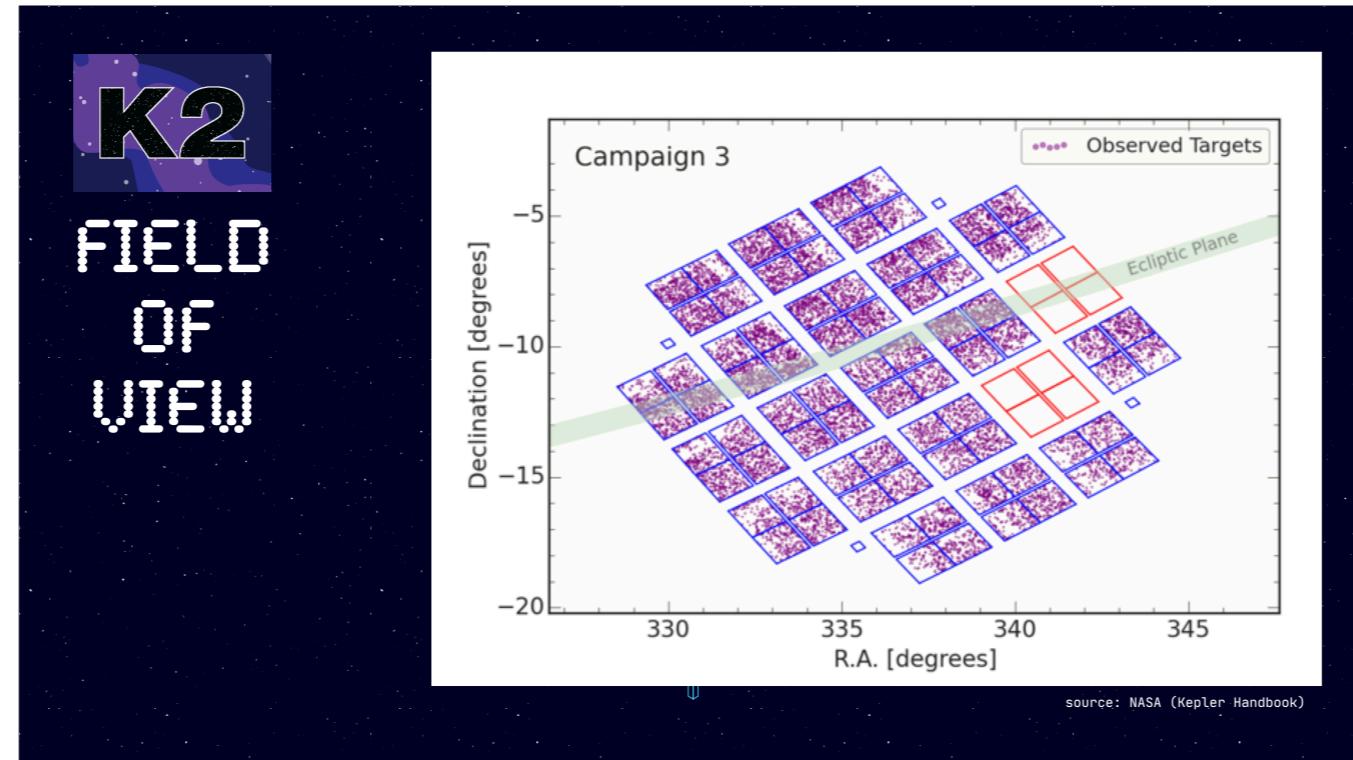
An exoplanet is a planet outside of our solar system, and that's what we're looking for at least in this first phase of the project.

Flux is a variation or change in light values of stars in this case. We will analyze flux values to make a prediction on whether or not a star may host a planet in its orbit.

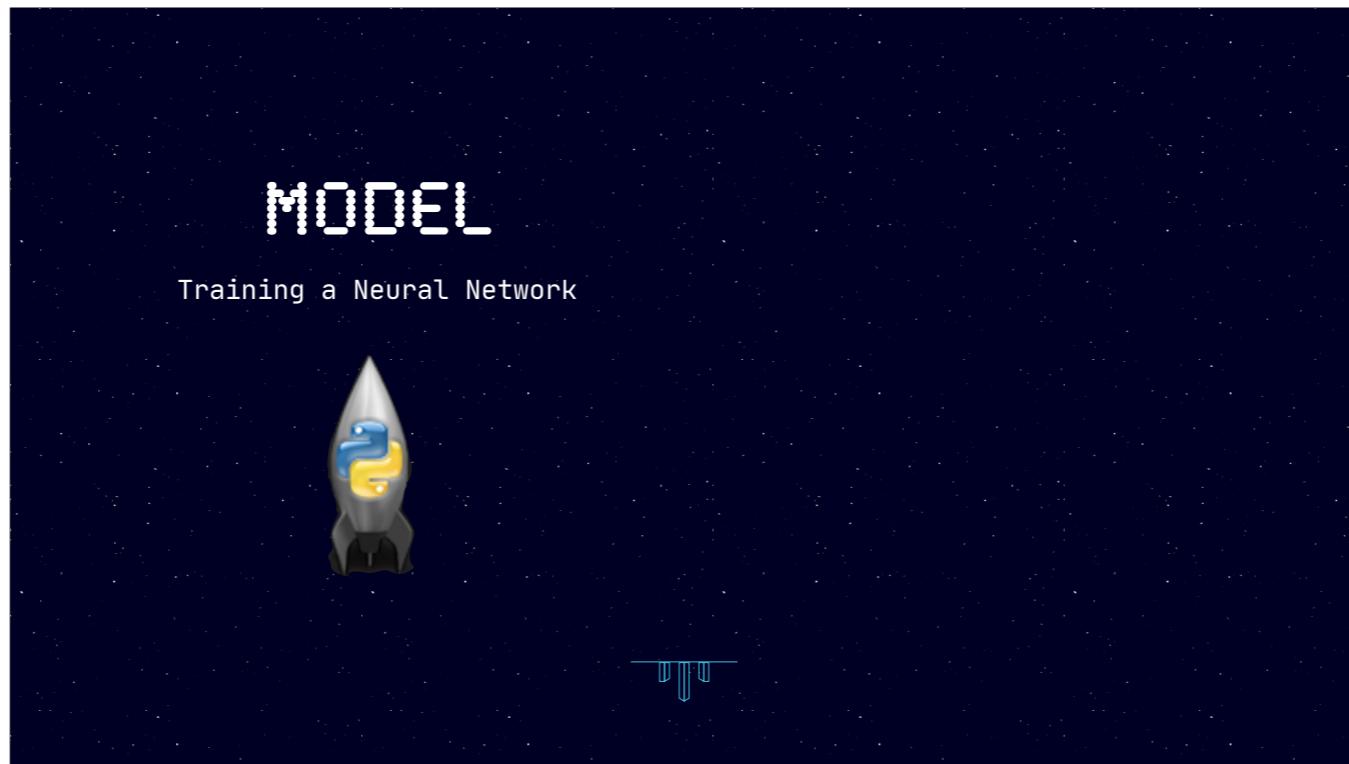
TCE or Threshold Crossing Event is what you see in the drawing there, the big yellow thing is a star, the red line represents level of brightness of the light emitted by the star, and when the black dot (which is a planet orbiting the star) crosses in front, it blocks some of the light coming toward us (if we're a telescope), so the values drop for a period of time, then go back up after the planet is no longer. So for this analysis, we're looking for that drop.



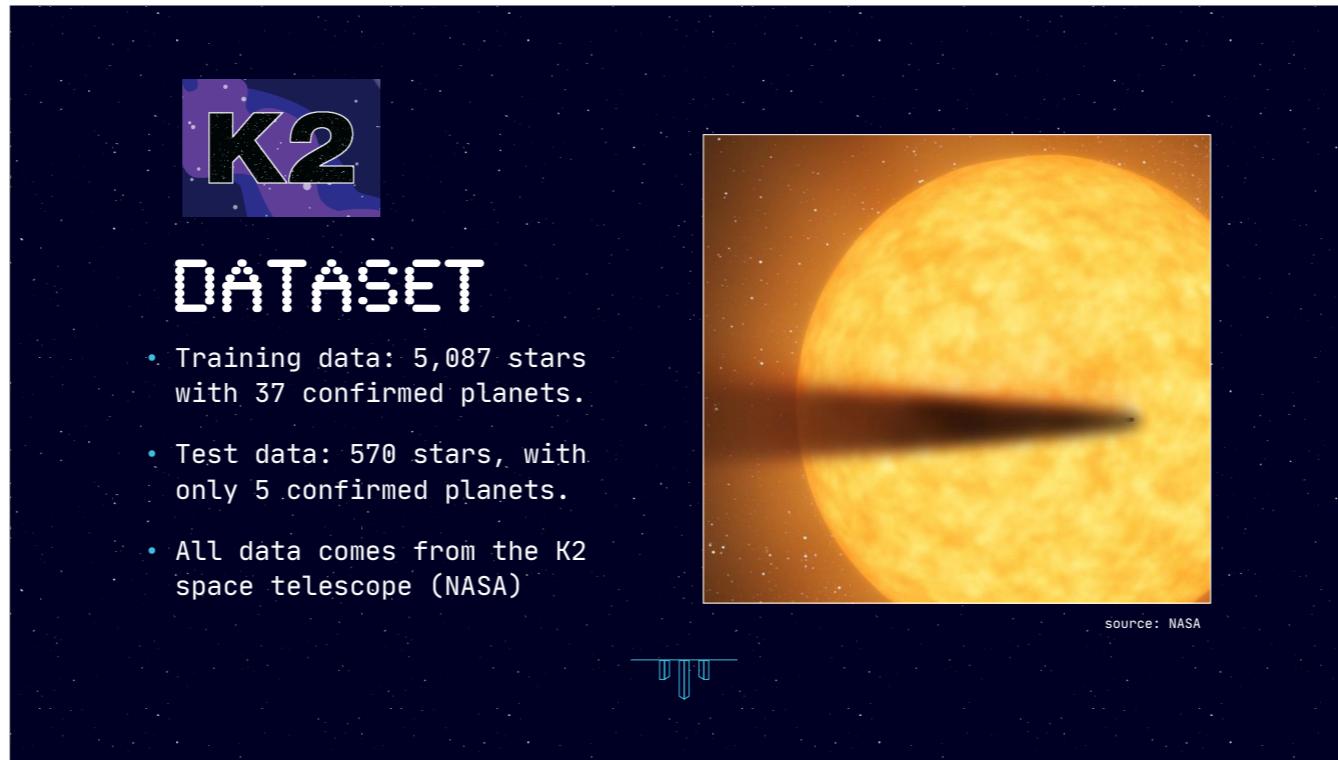
NASA's K2 mission included 20 campaigns, and for this initial phase of the project we're only looking at Campaign 3.



Campaign 3 includes flux values of stars in just one patch of sky over a period of time. Each campaign was supposed to be about 80 days before the telescope moved on to another of stars, however, just to make things more complicated, Campaign 3 was only around 60 days due to a technical issue in its data processing. I'd have to double check but I think it ran out of disk space, something I can relate to, on all my devices.



Training a neural network – this would be the brain of the artificial telescope, we are teaching it to identify which stars host planets, and which ones do not.



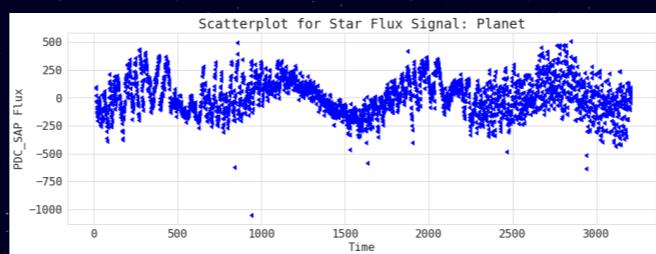
This artist's interpretation of a TCE is slightly more accurate than the other drawing, at least as far as scale goes. The training data including 3,197 flux observations for 5,087 stars: 37 of the stars have confirmed planets. The test data included 570 stars for testing the model: only 5 of the stars have confirmed planets.

So with 3,700 stars, our model needs to find just 42 confirmed planets. This means our data set is very sparse, which is actually a good thing because it means our model can help perform triage, filtering the most likely candidates of planet host stars, and dismissing the rest.

All data comes from the K2 (Kepler) space telescope (NASA)

TRAINING

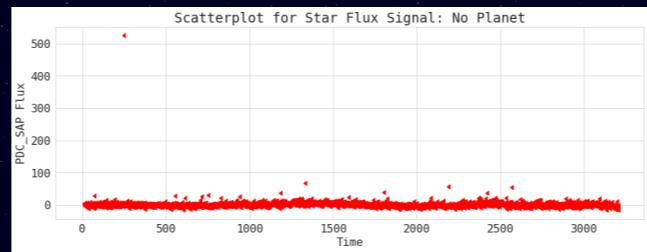
Looking for "dips" (periodicity) in the light flux of stars



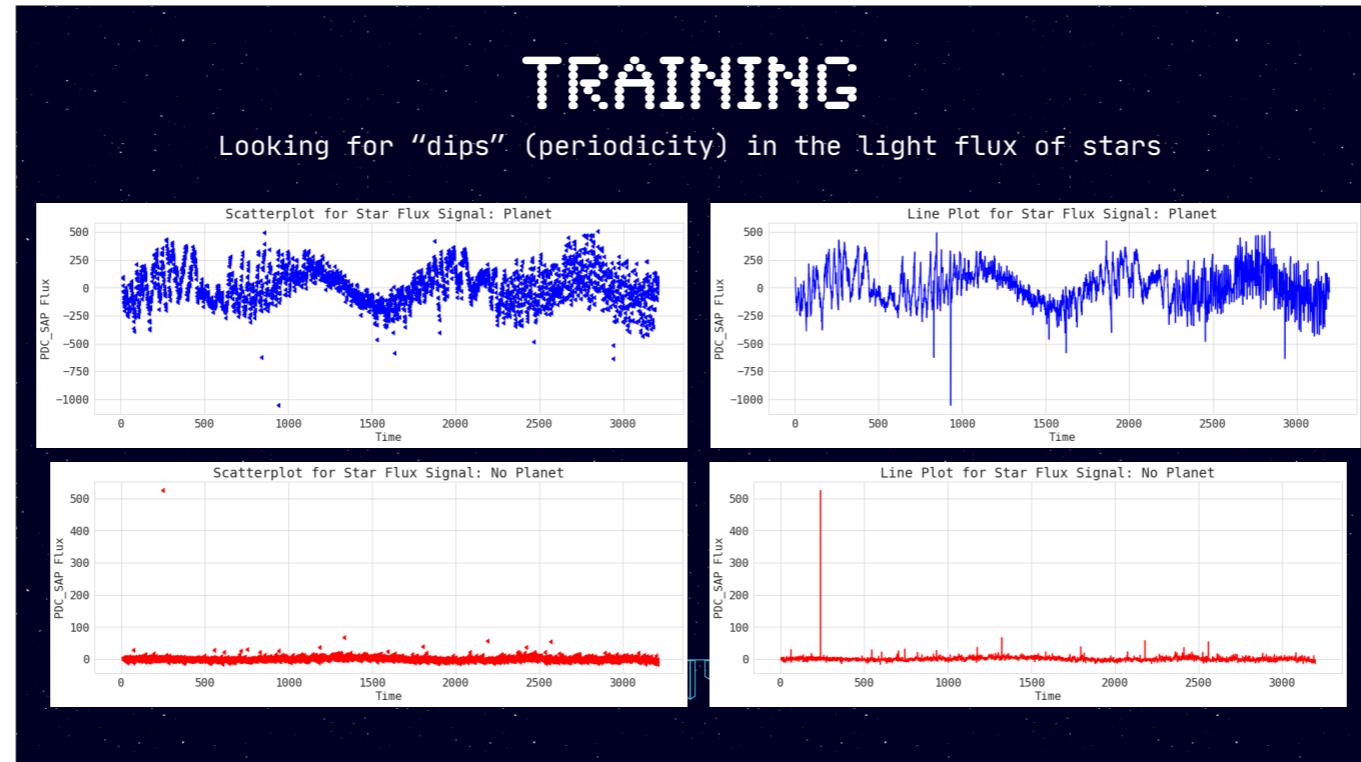
So we give our telescope the 60 days of flux values for about 5000 stars and we tell it hey this one has a planet...

TRAINING

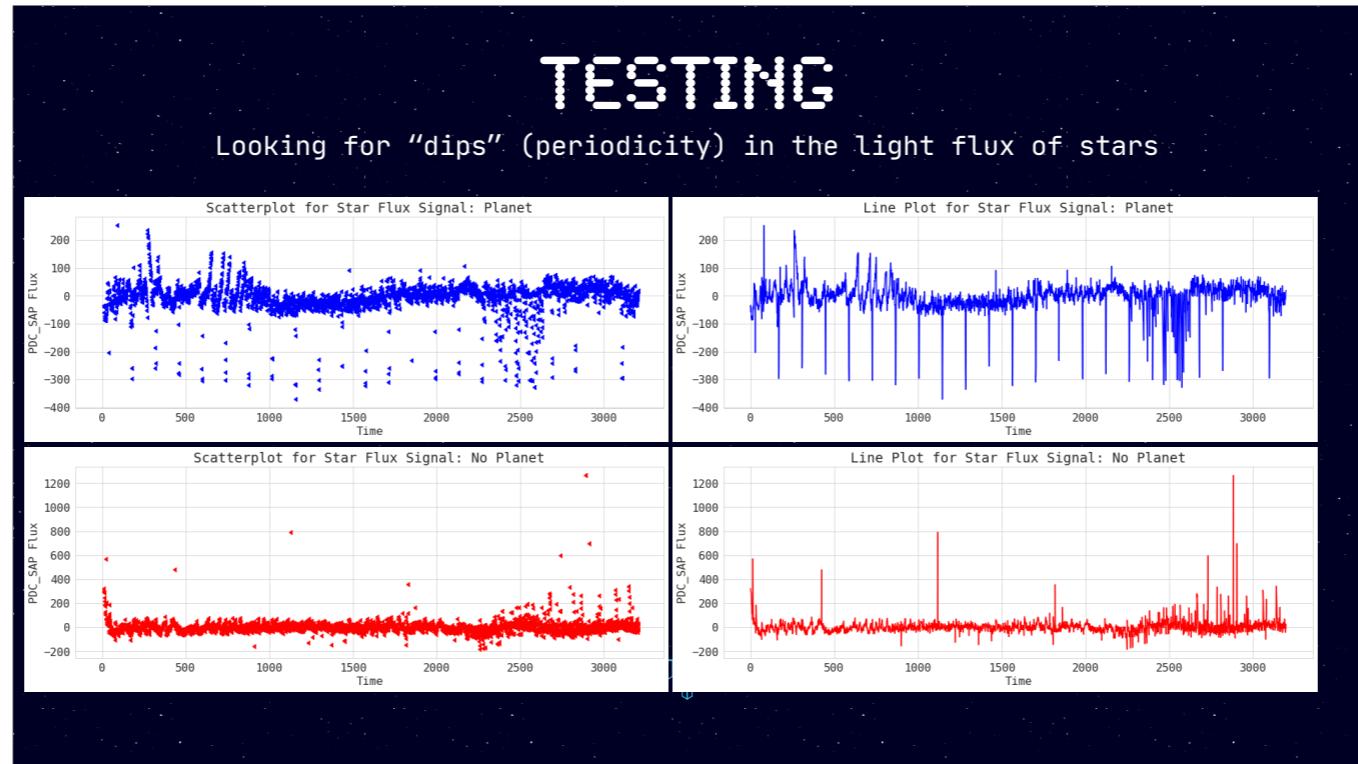
Looking for "dips" (periodicity) in the light flux of stars



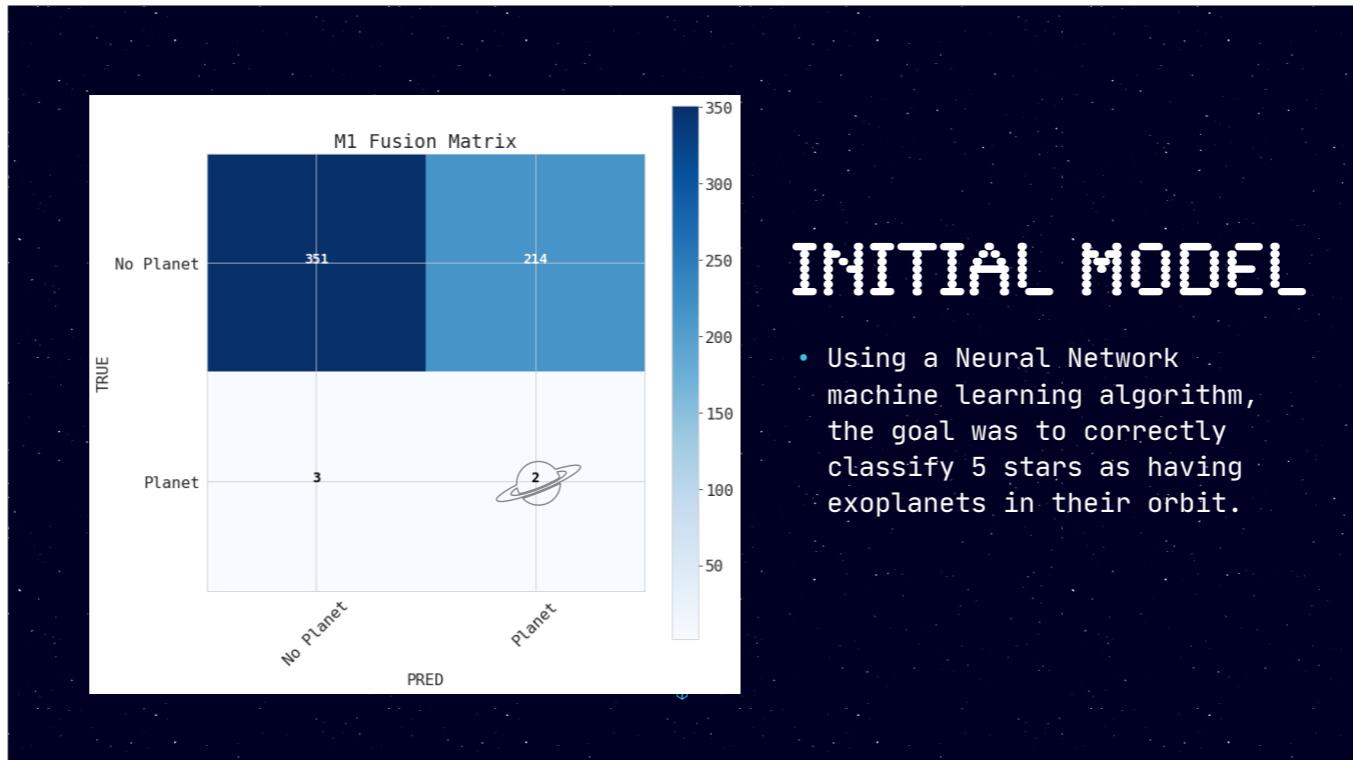
this one doesn't...



And do that over and over 5000 times. By the way this timeseries data originally didn't include any units for flux or time – I looked them up in the Kepler Handbook, crunched the numbers and used the Astropy and Astroquery libraries to create the correct timestamps and I'll be using those in the next phase of the project.



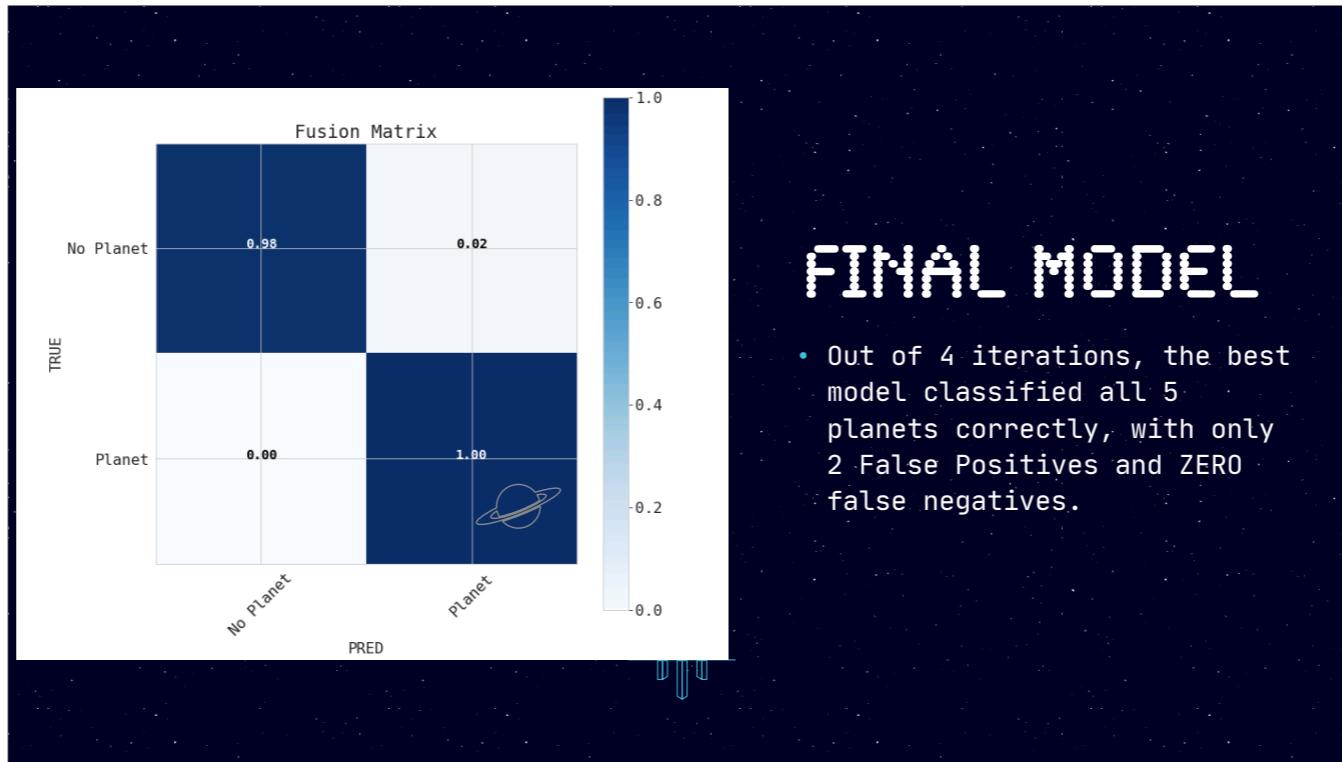
And then we take a smaller dataset of 570 stars, and we test it by asking it to make a prediction, including a probability percentage that it thinks this one has a planet, this one doesn't, then we asses how well the artificial telescope did...



INITIAL MODEL

- Using a Neural Network machine learning algorithm, the goal was to correctly classify 5 stars as having exoplanets in their orbit.

So again, this is a small data set to start out, training was ~ 5000 stars, test set was under 600. And at first it didn't do super well, it missed 3 of the 5 planet stars. So I made some adjustments to the learning process, and in the 2nd iteration...



FINAL MODEL

- Out of 4 iterations, the best model classified all 5 planets correctly, with only 2 False Positives and ZERO false negatives.

the model did extremely well. 100% accuracy, correctly classifying all 5 planets, and only mistook 2% of the nonplanet stars as planets. We call those false positives.

RESULTS

Accuracy: 99.64%

Recall: 100%

Jaccard Similarity: 71%

Fowlkes-Mallows: 99.64%



Some statistical measurements to mathematically assess the model's performance include accuracy and recall. Recall is basically how well does it learn from its mistakes, which is really key because that is essentially the definition learning, right? Jaccard and Fowlkes-mallows I included because these are "tougher" measures than accuracy, and are useful for sparse datasets – if most of the haystacks are empty, its easy for you do well guessing most of them are empty, so this kind of accounts for that problem.



1. Include stellar properties already calculated or possible to calculate using the MAST API
2. Use Fourier transform on the light kurves as part of the normalization scrubbing process.
3. Explore using computer vision on not only the Full Frame images we can collect from telescopes like TESS, but also on spectrographs of the flux values themselves.
4. Explore using autoencoded machine learning algorithms with Restricted Boltzmann Machines – this type of model has proven to be incredibly effective in the image analysis of handwriting as we've seen applied the MNIST dataset – let's find out if the same is true for images of stars or their spectrographs.



To continue this project, I'll take another approach for detecting exoplanets using computer vision to analyze images of spectrographs of this same data set then look at a larger dataset using an API with AWS. I will start with a Keras convolutional neural network as I did in this first phase, but ultimately the goal is to use Restricted Boltzmann Machines for each model, and then layer each model together to build a Deep Boltzman Machine that combines light curves, spectrographs, and full frame images from all three telescopes (K2, Kepler and TESS) wherever there is overlap in their campaigns.



The ultimate vision for this work will be to develop STARSKØPE into a front-end application that any astrophysicist can use to look at a single or collection of stars and have the model classify them according not only to exoplanet predictions, but also predict what type of star it is, and other key properties that would be of interest for astrophysical science applications.

THANK YOU!



APPENDIX

Images:

"NASA and ESA's past, current and future (or proposed) space missions with capacities to identify and characterize exoplanets." NASA / ESA / T. Wynne / JPL / Barbara Aulicino

"Kepler Field of View" from the Kepler Handbook

"Planet Bleed" NASA

"K2 Science" NASA

Screenshot from Stellarium Web Application

Documentation

Kepler Instrument Handbook and Supplement: <https://keplerscience.arc.nasa.gov/data/documentation/KSCI-19033-001.pdf>

K2 Handbook: http://archive.stsci.edu/k2/manuals/k2_handbook.pdf

Data

Exoplanet Hunting in Deep Space: <https://github.com/winterdelta/KeplerAI>