Project Nieuwestad: Modeling Exoplanet Habitability

Luke Meier, Ishaan Agrawal, and Aron Wang

Exoplanets

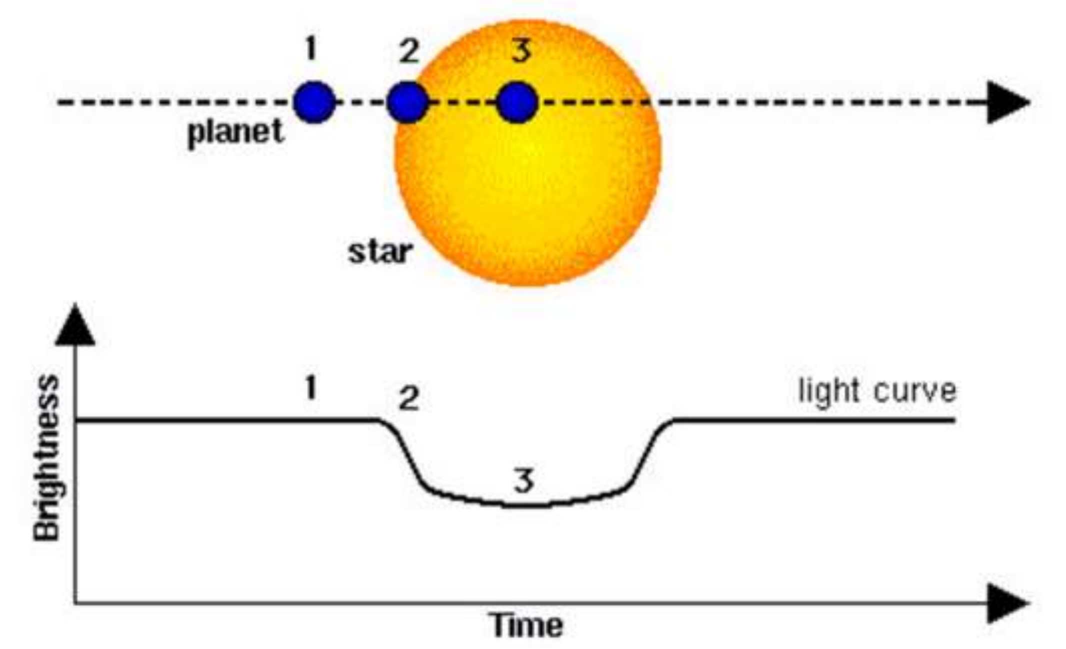
Astronomers estimate that there are hundreds of billions of planets in the Milky Way Galaxy. Those planets orbiting stars other than our own are called exoplanets, and some may have the capacity to host life. In March 2022, the number of confirmed exoplanet discoveries surpassed 5000.

Exoplanet Detection

Exoplanets are detected using several different methods. Doppler measurements, transit observations, microlensing, astrometry, and direct imaging are the five major methods, with transit observations being by far the most lucrative and popular. The majority of the thousands of confirmed exoplanets discovered with transit observations were found by NASA's Kepler Space Telescope, which was in operation from 2009 to 2018.

Transit Observations

Transit observations are the most successful method of exoplanet detection. Telescope point an instrument called a photometer at a star to observe the light it emits, and planets are found by the periodic dips in the stars brightness that they cause when their orbit crosses the star from the angle the telescope can see. While this is quantitatively the best method overall, it has limitations. Transit observations do a relatively poor job of detecting smaller, earth-sized planets.



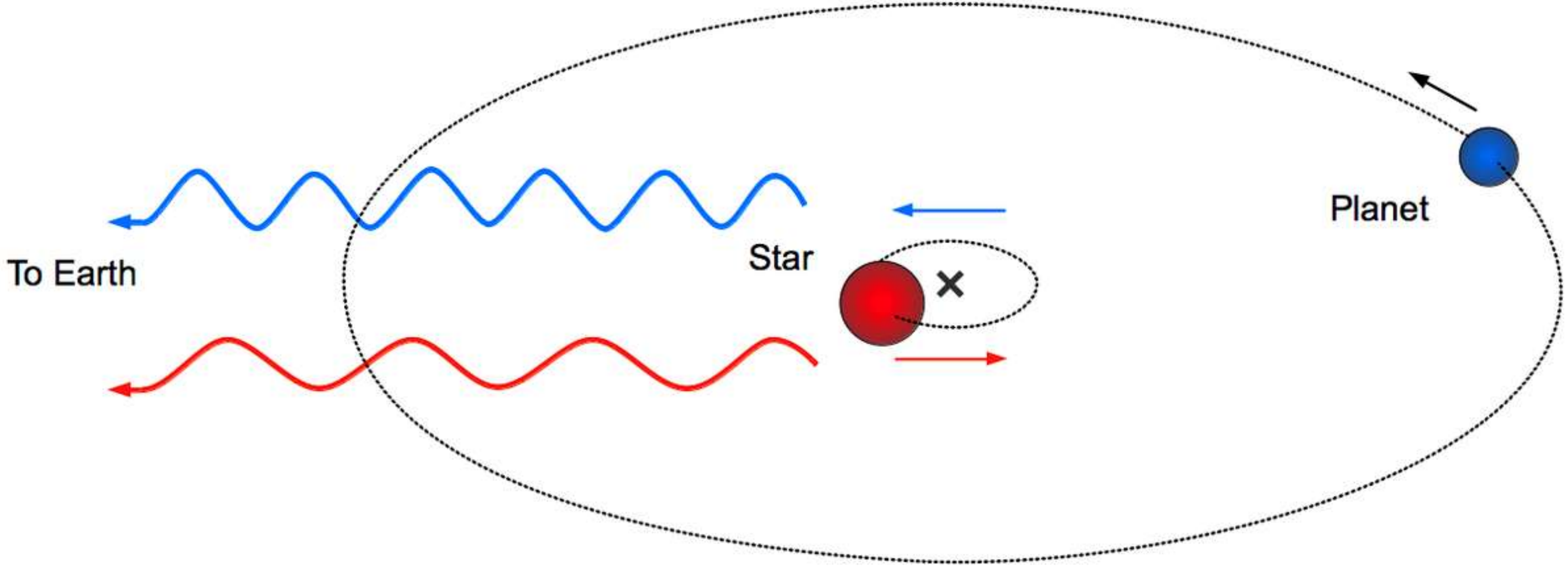
- ✗ The equation below represents the probability that a given planet can be detected by transit as a function of the planetary orbit, and the planetary and stellar radii.
- ✗ Omega is the angle of the planet's orbit, and e is the orbital eccentricity.

$$\mathcal{P}_{tr} = 0.0045 \left(\frac{\mathrm{AU}}{a}\right) \left(\frac{R_\star + R_\mathrm{p}}{R_\odot}\right) \left[\frac{1+e\cos(\pi/2-\omega)}{1-e^2}\right]$$

https://arxiv.org/abs/1505.0686
9

Radial Velocity

Before the rise of transit observations in the late 2000s, Radial Velocity was the most widely employed methodology by astronomers, where the small wobble in a star caused by an exoplanet's gravity is detected and catalogued. This wobble is detected by finding doppler shifts in the light from the star that a satellite detects. Radial Velocity has unique advantages and disadvantages: it can't detect planets further away, but it can detect planets of smaller mass.

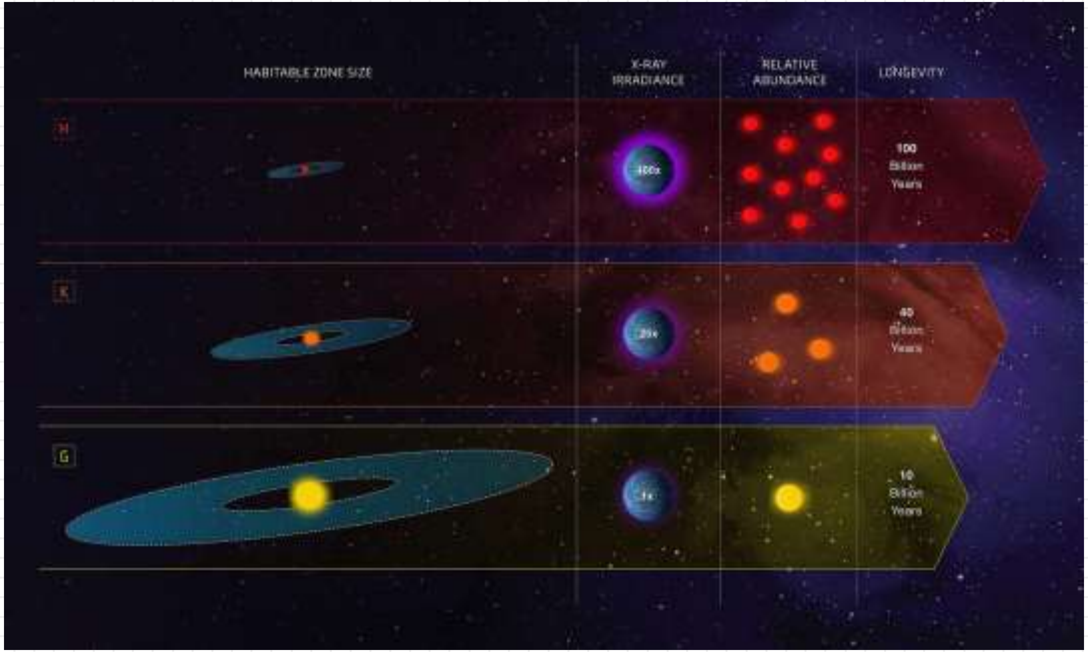


Research Question

What percentage of exoplanets are habitable?

Hypothesis

5-10% of discovered exoplanets will be likely candidates for habitability.



Stellar Properties

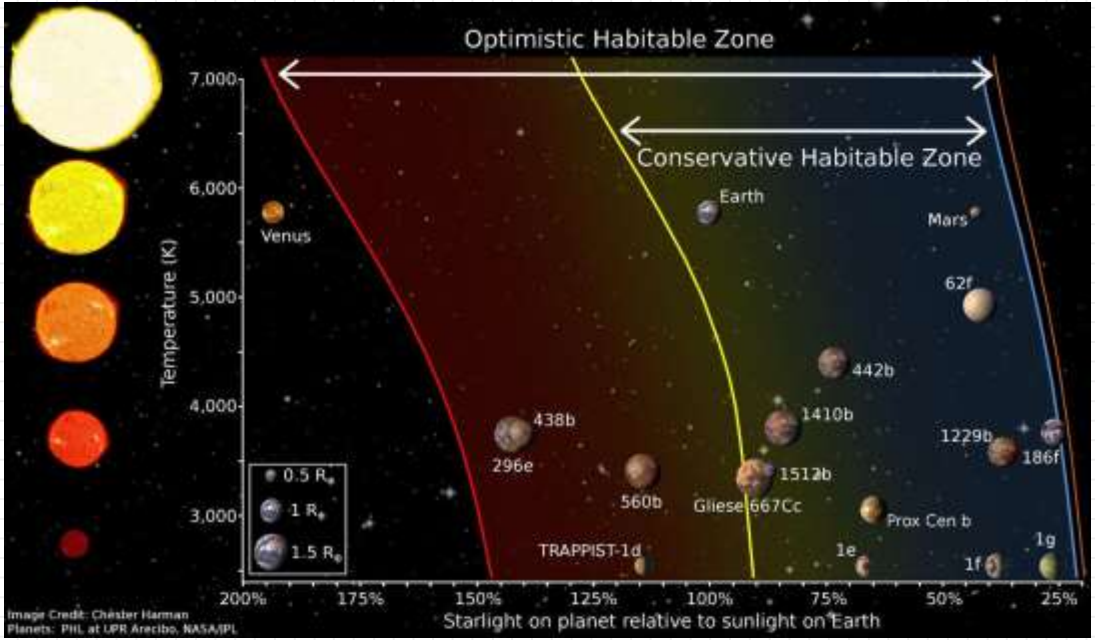
Stellar Properties: Habitability Ranges

- ✗ Type of stars:

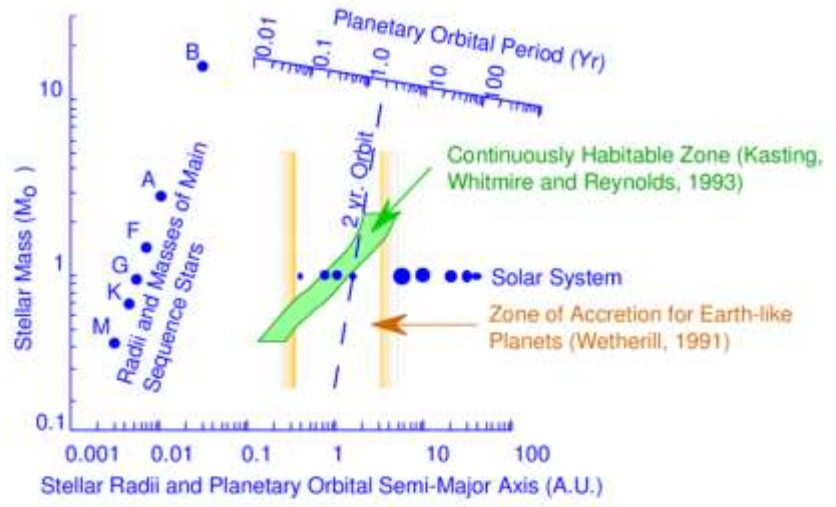
  K-type or G-type main sequence star
- ✗ Distance of the planet from the star:

  0.95-1.67 AU (Habitable Zone)
- ✗ Effective star surface Temperature:

  2,500-6,000K
- ✗ Stability of star
- ✗ Stellar radius
- ✗ # of Stars: 1 (no binary systems)
- ✗ Stellar surface gravity



Planetary Properties



- ✗ Magnetosphere
- ✗ Not tidally locked with star
- ✗ Tidally locked moon(s)
- ✗ Strength of magnetosphere relative to type and size of star
- ✗ Rotation speed of the planet: not too fast or too slow
- ✗ Orbital days
- ✗ Planet radius: between .75 and 3 earth radii

Temperature and Climate

- ✗ Orbital eccentricity below 0.1 for stable climate
- ✗ Must support liquid water
- ✗ Strong magnetic field
- ✗ Strong gravitational field (minimum of 2.7% the mass of earth to retain atmosphere)
- ✗ Must be in circumstellar habitable zone

The Data

NASA EXOPLANET ARCHIVE
NASA EXOPLANET SCIENCE INSTITUTE

Home | About Us | Data | Tools | Support | Login

Select Columns | Download Table | Plot Table | Download Data Products | View Documentation | User Preferences

Cumulative KOI Data

KepID	KOI Name	Kepler Name	Exoplanet Archive Disposition	Disposition Using Kepler Data	Disposition Score	Not Transit-Like False Positive Flag	Stellar Eclipse False Positive Flag	Centroid Offset False Positive Flag	Ephemeris Match Indicates Contamination False Positive Flag	Orbital Period [days]	Transit Epoch [BKJD]	Impact Parameter
			Confirmed									
10797460	K00752.01	Kepler-227 b	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	9.48803557±2.775e-05	170.53875±0.00216	0.146 +0.318 -0.146
10797460	K00752.02	Kepler-227 c	CONFIRMED	CANDIDATE	0.9690	0	0	0	0	54.4183827±0.000247	162.51384±0.00352	0.586 +0.059 -0.443
10854555	K00755.01	Kepler-664 b	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	2.525591777±3.761e-0	171.59555±0.00113	0.701 +0.235 -0.478
10872983	K00756.01	Kepler-228 d	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	11.09432054±2.036e-0	171.20116±0.00141	0.538 +0.03 -0.428
10872983	K00756.02	Kepler-228 c	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	4.13443512±1.046e-05	172.97937±0.0019	0.762 +0.139 -0.532
10872983	K00756.03	Kepler-228 b	CONFIRMED	CANDIDATE	0.9920	0	0	0	0	2.56658897±1.781e-05	179.55437±0.00461	0.755 +0.212 -0.523
10910878	K00757.01	Kepler-229 c	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	16.06864674±1.088e-0	173.621937±0.00051	0.052 +0.262 -0.052
11446443	K00001.01	Kepler-1 b	CONFIRMED	CANDIDATE	0.8110	0	0	0	0	2.470613377±2.7e-08	122.763305±8.7e-06	0.818±0.001
10666592	K00002.01	Kepler-2 b	CONFIRMED	CANDIDATE	1.0000	0	1	0	0	2.204735417±4.3e-08	121.3585417±1.6e-0	0.224 +0.159 -0.216
6922244	K00010.01	Kepler-8 b	CONFIRMED	CANDIDATE	0.9980	0	0	0	0	3.522498429±1.98e-07	121.1194228±4.71e-0	0.631±0.007
10984090	K00112.02	Kepler-466 c	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	3.709214104±6.536e-0	133.98318±0.00143	0.051 +0.395 -0.051
10526549	K00746.01	Kepler-660 b	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	9.27358173±1.037e-05	173.258155±0.00087	0.387 +0.004 -0.386
10583066	K00747.01	Kepler-661 b	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	6.02930329±5.509e-06	171.602959±0.00071	0.258 +0.196 -0.258
10601284	K00749.01	Kepler-226 c	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	5.349553819±8.834e-0	171.80694±0.00127	0.092 +0.369 -0.092
10601284	K00749.02	Kepler-226 b	CONFIRMED	CANDIDATE	0.9800	0	0	0	0	3.94105221±1.094e-05	136.08662±0.00233	0.226 +0.243 -0.226
10601284	K00749.03	Kepler-226 d	CONFIRMED	CANDIDATE	0.9710	0	0	0	0	8.10904807±6.787e-05	132.8031±0.00716	0.025 +0.425 -0.025
10662202	K00750.01	Kepler-662 b	CONFIRMED	CANDIDATE	1.0000	0	0	0	0	21.67699193±7.01e-05	171.53594±0.00258	0.896 +0.01

Showing records 1 to 18 of 2669 (9564 total) DOI 10.26133/NEA4

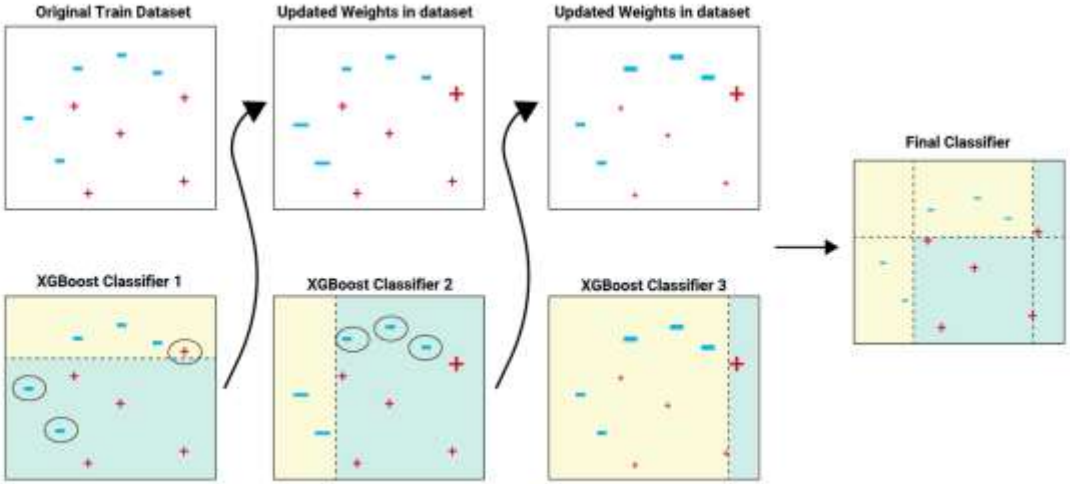
Clear Checked | Check All | Show Filters | Reset Filters

Artificial Intelligence Methodology

- We created an **Artificial Intelligence forecasting model** that analyzes the parameters of exoplanets and rates them relative to the ideal habitability conditions for life to exist. We inputted exoplanet data into our model to calculate the likelihood of each planet to host life.

Methodology for AI Learning Process:

1. Preprocess/filter data
2. Removing absent/irrelevant data
3. Look for correlations and perform data analysis
4. Prepare data for the machine learning model
5. Define/program model
6. Train the model
7. Assess the model
8. Refine the model



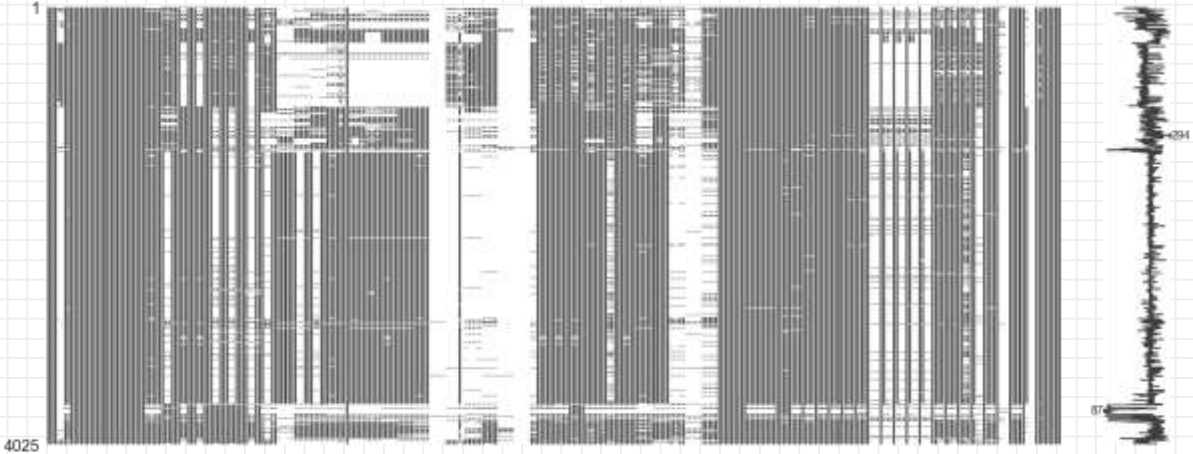
Artificial Intelligence Methodology

We built the artificial intelligence model by importing open source machine learning libraries and constructing the model in the python programming language. The model is developed using the 8 standard steps of machine learning.

Machine Learning Flowchart

Pre-Processing

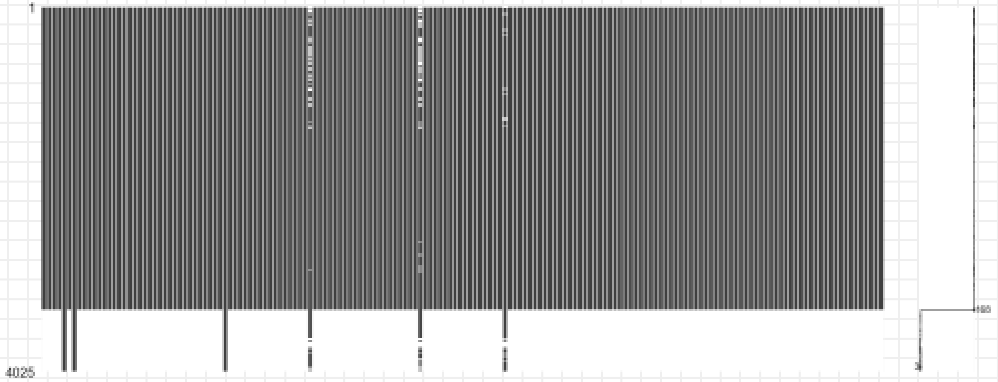
Before: Lots of missing data, hard to use



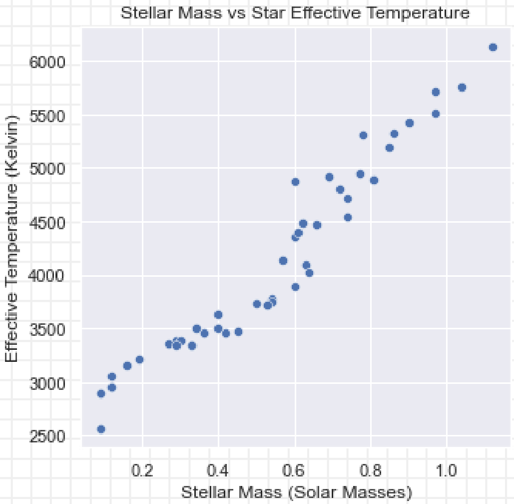
Data Pre-Processing

- ✗ We started by gathering confirmed exoplanet data from NASA and PHL.
- ✗ Astronomical data tends to contain lots of missing values, which is bad for the machine learning algorithm.
- ✗ To use the data, we have to clear out the null values and fill the table with usable data.
- ✗ We decided to remove all of the columns where over 40% of the data was missing, as it would not be easy to calculate values for the machine learning model.
- ✗ For the remaining null values, we calculated them using a mathematical system that generates the values based on the values of every other planet and every other cell in the dataset.

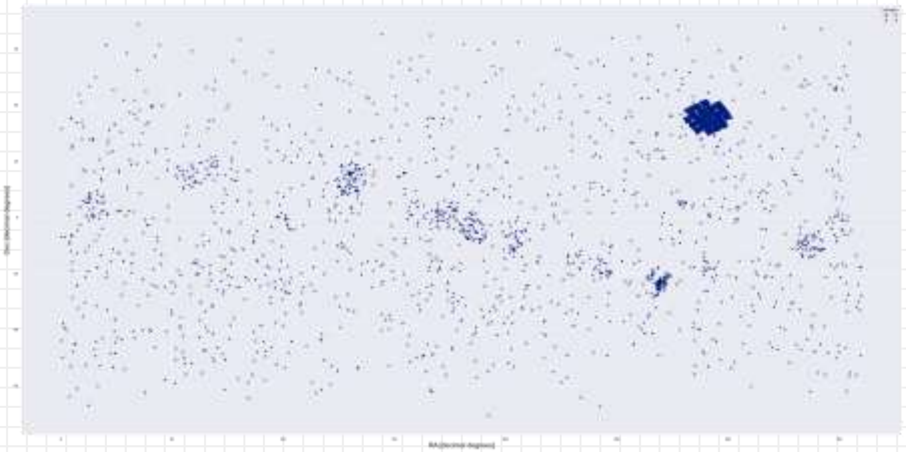
After:
Less missing data, easier to use



Finding Preliminary Correlations



To start our research, we looked into some of the correlations between the various parameters of planets and stars.



Average Stellar Mass for the host star of a habitable exoplanet: 0.4978 solar masses
Average effective surface temperature of the host star of a habitable exoplanet: 3979.8 K

Stellar Metallicity: We found that stars with habitable planets tend to have a much higher amount of metal, compared to stars that do not.

The Goal of The Model

What It Does

- In our model construction, we trained a machine learning algorithm to identify whether or not a given exoplanet is habitable by giving it a database with exoplanet features, where exoplanets were already labeled habitable or not-habitable.
- The machine learning model analyzed various features of the stellar systems and their exoplanets to determine whether or not a given planet is habitable.
- The model calculated how the features of a stellar system and the planets within it relate to whether or not the system contains any habitable planets.

Model Performance Assessment

Training The Correct Model

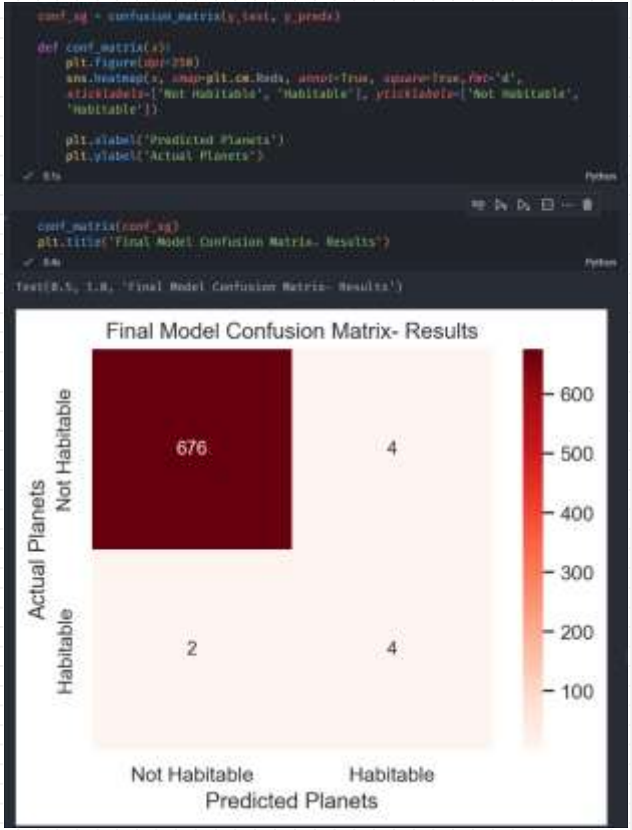
- We will train multiple different algorithms and measure the accuracy of the various algorithms, and use the algorithm with the highest F1 score (a combination of precision and recall) to predict the habitability of exoplanets.
- We trained 9 different models and measured the accuracy, precision, recall, and F1 scores of all of the models.
- XGBoost had the highest F1 score, showing that it was the most balanced model overall.

Model	Train Accuracy	Test Accuracy	Precision	Recall	F1
K-Nearest Neighbors	98.13	94.90	0.56	0.89	0.60
Logistic Regression	95.47	92.86	0.55	0.88	0.57
XGBoost	99.61	99.13	0.75	0.83	0.78
Gaussian Naive Bayes	91.68	96.79	0.57	0.74	0.60
Decision Trees	100.00	98.98	0.70	0.66	0.68
Random Forest	100.00	98.83	0.66	0.66	0.66
Stochastic Gradient Boosting Classifier	99.36	98.69	0.64	0.66	0.65
AdaBoost Classifier	100.00	98.83	0.66	0.66	0.66
Hard Voting	99.38	98.54	0.58	0.58	0.58

Model Performance Assessment

XGBoost Performance Assessment

- ✗ We chose to use the XGBoost machine learning algorithm to process our data.
- ✗ We trained the algorithm using our training dataset, and then tested its accuracy, generating this confusion matrix.
- ✗ We can see that the model was very accurate at identifying the non-habitable planets, correctly identifying all of them. It also correctly identified some habitable planets, but had a few false positives.



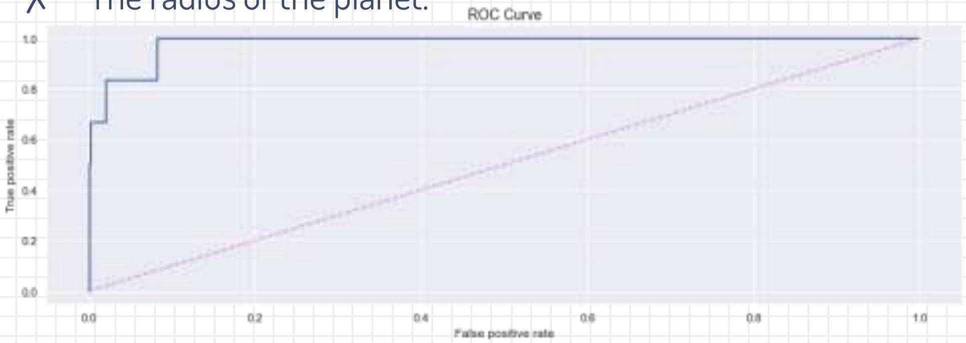
Model Results

Analyzing Features

- Some of the models we trained created built a list of parameters corresponding to which features are most important to detect the habitability of a planet.
- By looking at the features which were determined to be most important to making a planet habitable, we gained valuable insights into the nature of habitable star systems, their basic design, and the critical elements of them.

Most Important Features for Planet Habitability

- The most important features of a habitable planet were:
  - The planet's insolation flux
  - The planet's effective temperature
  - The ratio between the radius of the star and the apogee of the planet's orbit.
  - The radius of the planet.



feature	importance
pl_insol	0.209611
pl_eqt	0.179109
pl_ratdor	0.086717
pl_rade	0.063620
pl_insolerr1	0.055801
pl_radj	0.045446
st_teff	0.040490
st_logg	0.033707
pl_orbper	0.029043
st_mass	0.021580
sy_pmraerr1	0.016024
st_agelim	0.015707
sy_gaiamagerr2	0.013552
sy_gaiamagerr1	0.012845
st_loggerr1	0.011792
pl_radeerr2	0.011029

Conclusions

Hypothesis Rejection

- ✗ We predicted that 5-10% of all exoplanets would be labeled habitable.
- ✗ We discovered that 2.67% of all catalogued exoplanets are predicted habitable.
- ✗ While K and G-type stars are most likely to host life, a variety of other factors also play a role in whether or not a planet can host life, which narrows down the possibility of life significantly.

Habitable Planets

- ✗ Out of 5,063 confirmed exoplanets, our model predicted that 135 of them are habitable.

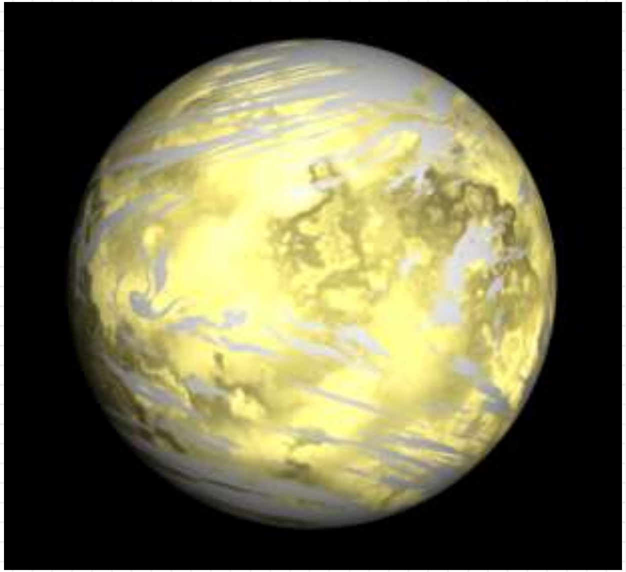
Planet Parameters

- ✗ The size of the star is the most critical factor in determining whether or not the planets around it will be habitable.
- ✗ The radius and the orbital period of the planet also play a major role.

Conclusions

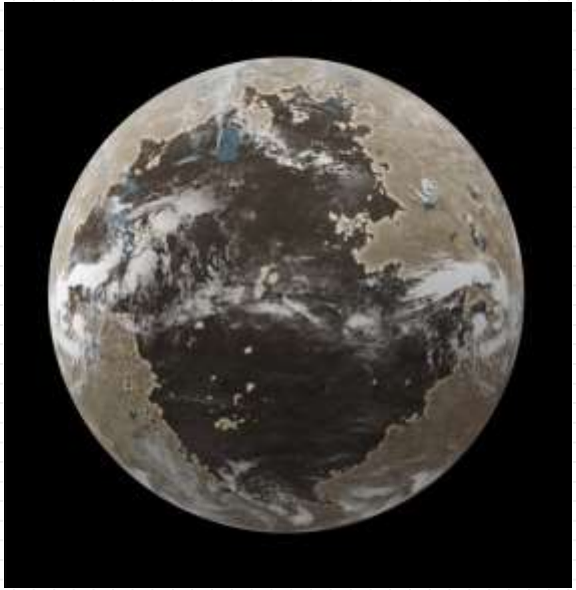
Farthest Exoplanet from Earth Labeled Habitable by our Model

- ✗ Kepler 1638b
- ✗ Distance from Earth: 1525.52 parsecs



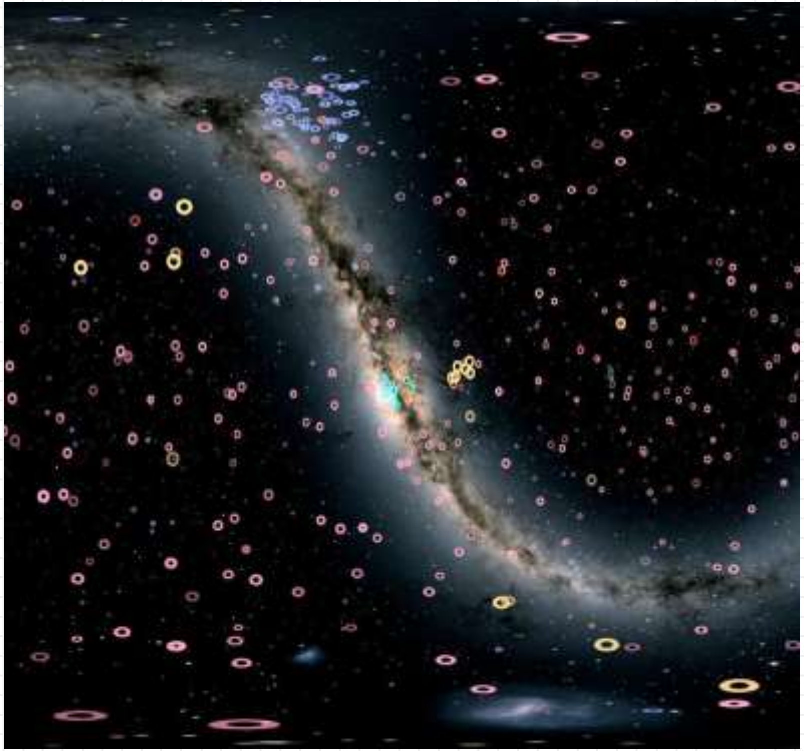
Nearest Exoplanet to Earth Labeled Habitable by our Model

- ✗ Proxima Centauri b
- ✗ Distance from Earth: 1.301 parsecs



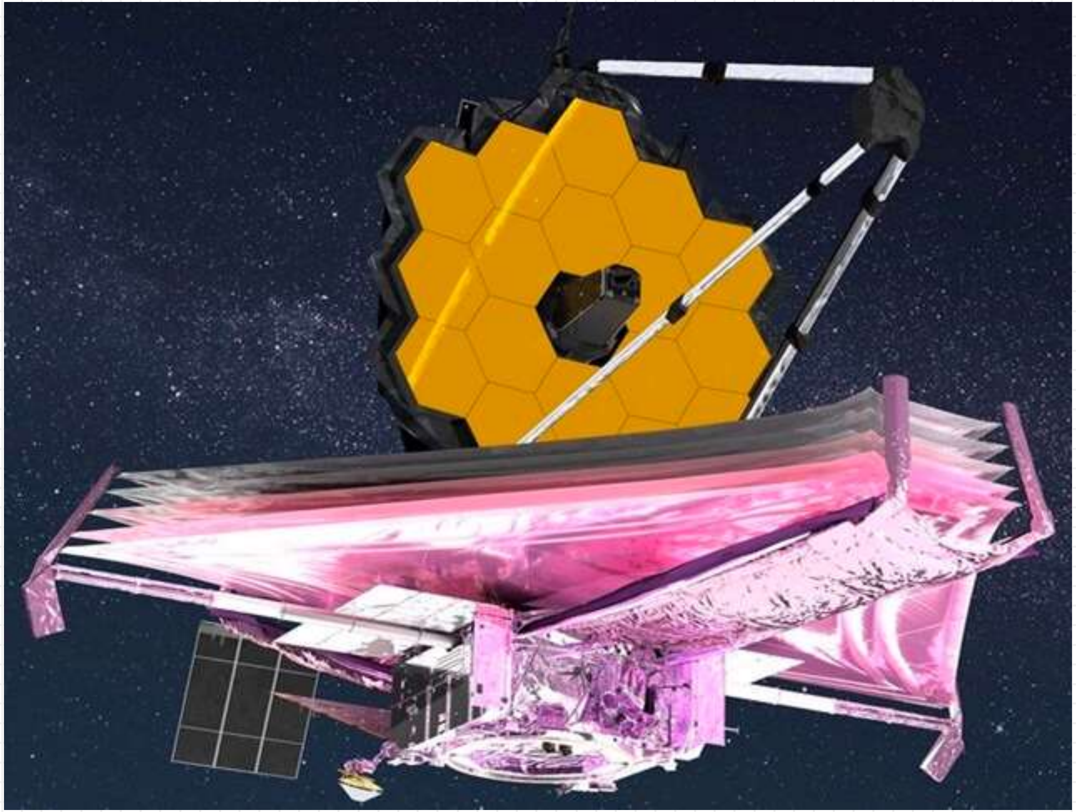
Application in Exoplanetology

- ✗ The study of exoplanetology is a field that will continue to grow in the coming years. It is a subfield of astronomy that deals with the techniques of discovering extrasolar planets and the analysis of the dynamic processes of exoplanets, planetary systems, and the properties of a planet's host star.
- ✗ Exoplanetology an interdisciplinary field that integrates the studying of astrobiology, astrophysics, astronomy, astrochemistry, astrogeology, geochemistry, planetary science, and exoplanet statistics.
- ✗ Our research is in the subfield of exoplanet statistics, which draws astronomical conclusions from the analysis of exoplanet databases.
- ✗ Astrobiology, another distinct subset of exoplanetology, searches for possible life on other planets and in other solar systems while also investigating how early life forms. Our research provides this field with an estimate of how common habitability is in our galaxy.



Map of confirmed exoplanet locations in the Milky Way.

The Future of Exoplanet Detection



- The James Webb Space Telescope is an infrared space telescope that will also use transit observations to discover new exoplanets further from earth than ever have been found before.
- Using spectroscopy, JWST will be able to identify what elements are present in an exoplanet's atmosphere. As this new data becomes available, exoplanet statistics in the context of habitability will have the capacity to become much more precise as models such as ours will have more parameters to use in calculations.

Image Citations

- https://www.youtube.com/watch?v=4F7sdy2rZws
- https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=PS
- https://exoplanets.nasa.gov/what-is-an-exoplanet/overview/
- https://exoplanets.nasa.gov/
- https://earthsky.org/space/goldilocks-stars-g-k-dwarfs-best-for-alien-life/
- https://phl.upr.edu/
- https://www.ipac.caltech.edu/project/exoplanet-archive
- https://en.wikipedia.org/wiki/Circumstellar_habitable_zone
- https://exoplanets.nasa.gov/search-for-life/habitable-zone/
- https://www.universetoday.com/24299/types-of-stars/
- https://www.researchgate.net/figure/Relationship-of-Stellar-Properties-to-the-Habitable-Zone-and-Kepler-Search-Space-Each_fig1_23896524
- https://www.bostonglobe.com/metro/2016/07/20/mit-scientists-say-planets-orbiting-dwarf-star-have-atmospheres-that-may-support-life/cxg3vcQZupT5ImAD16Pj4J/story.html
- https://www.bbc.com/news/science-environment-27941557
- https://www.nasa.gov/mission_pages/sunearth/news/storms-on-sun.html
- https://chem.libretexts.org/Bookshelves/General_Chemistry/Book%3A_General_Chemistry%3A_Principles_Patterns_and_Applications_(Averill)/01%3A_Introduction_to_Chemistry/1.09%3A_Essential_Elements_for_Life
- https://www.geeksforgeeks.org/flowchart-for-basic-machine-learning-models/
- https://www.youtube.com/watch?v=aiFD_LBx2nM