Exoplanet Prediction

# **David Kinney Spring 2021** [dkinneyBU/DSC680 (github.com)](https://github.com/dkinneyBU/DSC680)

# Domain

The NASA Exoplanet Archive is an online astronomical exoplanet and stellar catalog and data service that collates and cross-correlates astronomical data and information on exoplanets and their host stars. These data include stellar parameters (such as positions, magnitudes, and temperatures), exoplanet parameters (such as masses and orbital parameters) and discovery/characterization data (such as published radial velocity curves, photometric light curves, images, and spectra). [1]

References

[1] NASA Exoplanet Archive – NASA Exoplanet Science Institute

<https://exoplanetarchive.ipac.caltech.edu/index.html>

[2] [What Is an Exoplanet? | NASA Space Place – NASA Science for Kids](https://spaceplace.nasa.gov/all-about-exoplanets/en/)

[3] [Overview | What is an Exoplanet? – Exoplanet Exploration: Planets Beyond our Solar System (nasa.gov)](https://exoplanets.nasa.gov/what-is-an-exoplanet/overview/)

Planet Hunters TESS

[Planet Hunters TESS | Zooniverse - People-powered research](https://www.zooniverse.org/projects/nora-dot-eisner/planet-hunters-tess)

[What in the World is an ‘Exoplanet?’ | NASA](https://www.nasa.gov/feature/jpl/what-in-the-world-is-an-exoplanet)

”Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow”

1. Geron 2019 O’Reilly Media Sebastopol, CA

# Data

The dataset I am leveraging is the **Kepler Object of Interest** (KOI) table, located on the NASA Exoplanet Archive website noted above. The **Data Columns** documentation can be reviewed here: [Data columns in Kepler Objects of Interest Table (caltech.edu)](https://exoplanetarchive.ipac.caltech.edu/docs/API_kepcandidate_columns.html). The dataset can be found here: [Kepler Objects of Interest (caltech.edu)](https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative). The dataset consists of 9,565 rows with 2,358 confirmed exoplanets, 2,366 candidates and 4,840 false positives.

# Research Questions and Benefits

* Can I identify the features within the dataset that are “better” at predicting whether an observed planet is an exoplanet?
* Are certain features more attuned to catching false positives?
* Does the size of a planet affect its ability to sustain carbon-based life (temperature, surface gravity, etc.)?
* While there is a large focus on transit measurements, what other observations affect the identification of an exoplanet—a light curve inconsistent with an exoplanet, the existence of an eclipsing binary star, etc.?

# The “Why”

Why is this research important? Of the literally hundreds of exoplanets discovered to date, not one is reachable by current technology. But Man has always looked to the stars, wondering if we are alone. By narrowing down the trillions upon trillions of bodies orbiting stars to exoplanets, astronomers can then measure the size, distance from the star, and temperature of the exoplanet towards the end goal of calculating the chances that these worlds can support carbon-based life.

# Method

I intend to initially employ a Decision Tree algorithm; depending on the results, it is likely I will fit a Random Forest algorithm as well. In addition, it is possible that as a pre-processing step I will leverage Principal Component Analysis (PCA) for the purpose of dimensionality reduction. The KOI dataset is rich in features and therefore conducive towards predicting whether a candidate astral body is an exoplanet (or not). However, quantity of features is both a blessing and a curse, so as mentioned I will also apply dimensionality reduction and will decide if it improves model performance.

# Potential Issues?

A major challenge for me throughout the Data Science program, and Machine/Deep Learning in general, is taking an intelligent approach to model selection. I am often off the mark as to which model type is best applied to which analysis scenario. I *think* that leveraging Decision Trees and Random Forests makes sense, but once I get further into the project, I may discover I have taken a wrong turn. If that turns out to be the case, I need to course-correct quickly in order to stay on track to meet deadlines for all my deliverables.

# Concluding Remarks

“Whether life exists beyond Earth is one of the most profound questions of all time. The answer will change us forever, whether it reveals a universe rich with life, one in which life is rare and fragile, or even a universe in which we can find no other life at all.” [3]

Exoplanets are planets that orbit around a star, as in our solar system. These bodies are very hard to see directly with telescopes since they are hidden by the bright glare of the stars they orbit. The Kepler Objects of Interest (KOI) dataset contains various measurements of transit, in addition to many others that aid in identifying exoplanets. In 2009, NASA launched the **Kepler** spacecraft to search for exoplanets. Kepler looked for planets in a wide range of sizes and orbits that circled around stars of varied size and temperature. [2] The Kepler spacecraft detected exoplanets using the **transit** method. As a planet passes (transits) in front of a star, it blocks out a bit of the star’s light. By observing the stars’ change in brightness astronomers can figure out the size of the orbiting planet as well as how far away it is from the star. Further, this data then aids in calculating the planet’s temperature, and the chances that it may contain liquid water—the stuff of life…

Why expend all this effort in time and resources to discover potentially habitable planets, when we fully realize we can never visit even one of them? One simple answer is, *because we can.* It seems that humans, by their very nature, are explorers, unable to sit idly by while a vast, mysteriously beautiful universe unfolds before us. When we look up at the night sky, we are alternately comforted by its splendor, and saddened while we wonder, *are we alone in all this vastness?* Perhaps searching for exoplanets is our way—possibly our only way—of ever concluding that there is at least a chance that we are *not* alone…