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, and provides tools to work with these data. The archive is dedicated to collecting and serving important public data sets involved in the search for and characterization of extrasolar planets 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>

Planet Hunters TESS

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

[Asteroid Database and Mining Rankings - Asterank](http://www.asterank.com/)

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

[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 table, located on the NASA Exoplanet Archive 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? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.

Exoplanets are planets that orbit around a star, as in our solar system. These bodies are very hard to see directly with telescopes due to the fact that they are hidden by the bright glare of the stars they orbit. Kepler detected exoplanets using something called the transit method. When a planet passes in front of its star, it’s called a transit. As the planet transits in front of the star, it blocks out a little bit of the star's light. That means a star will look a little less bright when the planet passes in front of it. Astronomers can observe how the brightness of the star changes during a transit. This can help them figure out the size of the planet. So, astronomers use other ways to detect and study these distant planets. They search for exoplanets by looking at the effects these planets have on the stars they orbit. By studying the time between transits, astronomers can also find out how far away the planet is from its star. This tells us something about the planet’s temperature. If a planet is just the right temperature, it could contain liquid water—an important ingredient for life.

The Kepler Objects of Interest (KOI) dataset contains various measurements of transit, in addition to many others that aid in identifying exoplanets.

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 at the stars, wondering if we’re 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 endo goal of calculating the chances that these worlds can support carbon-based life.

# Method

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

I intend to initially employ a Decision Tree algorithm; depending on the results, it’s likely I will fit a Random Forest algorithm as well. In addition, it’s possible that as a pre-processing step I will leverage Principle Component Analysis (PCA) for the purpose of dimensionality reduction.

# Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

# Concluding Remarks

Tie it all together. Think of this section as your final report’s abstract.