Project McNulty Proposal
Classifying an Object of Interest as an Exoplanet or not

Luke Tibbott

Overview

The astronomical community has a rigorous process for deciding whether something is an exoplanet or not. Before an object is classified as an exoplanet, it is an "object of interest." Many of these go on to be false positives, meaning there was a measurement error, the object is something other than an exoplanet, or a multitude of other possibilities. My aim is to examine this dataset, and use a classification algorithm to determine whether an object of interest is an exoplanet or not.

I found my dataset on Kaggle, which has information on 10,000 of these "objects of interest." The data was provided by NASA. Each object is classified as either **Candidate, Confirmed,** or **False-Positive.** Objects that are **candidates** will be excluded from my analysis, as a determination has not yet been made on these objects of interest.

After cleaning and dropping, there are 2,200 **confirmed** and 3,900 **False-positive** observations. This is a total of about 6,100 observations.

The dataset includes 50 features, but many of these are not of interest for my analysis.

I have narrowed my interest to the following features (please note all features are continuous):

Feature	Description
Disposition	Confirmed exoplanet or False-positive
Period	Interval between consecutive planetary transits
Time of discovery	Time corresponding to center of the first detected transit in Barycentric Julian Day minus 2,454,883 days (corresponds to 1-1-2009)
Impact	Sky-projected distance between center of stellar disc and center of the planet disc at conjunction, normalized by stellar radius
Duration	Duration of observed transits

Depth	Fraction of stellar flux lost at the minimum of planetary transit
Radius	Radius of the object
Planetary Temperature	Approximate temperature of planet
Insolation	Temperature relative to Earth
Model SNR	Transit depth normalized by the mean uncertainty in the flux during transits
Planet number	Number this planet has in its solar system (eg Earth would be 3)
Star Temperature	Photospheric temperature of host star
Gravity	Base-10 logarithm of the acceleration due to gravity at the surface of the *star*
Right Ascension	East-West measurement of where the star is in Earth's night sky
Declination	North-South measurement of where the star is in Earth's night sky
Kepler-band magnitude	Magnitude of the star (how bright the star is)

Known unknowns

I am not an astronomer, but I have studied astronomy. I don't know what all these features mean, so one of my first steps is going to be to understand what everything means. That being said, feature engineering might be slightly more challenging for this project given the highly domain-specific information in this dataset

I'm also not sure of the predictive power of this dataset. It's possible that this is a really hard problem in astronomy that no one has found a good solution to. If every object of interest was easy to classify as not an exoplanet, it wouldn't make it on to the list of objects of interest. This makes me think that the objects on this list are at least very convincing impostors, if not indistinguishable.

That being said, I'm going to try my best to get an accurate classification model and I think it'd be really cool if my model was able to correctly predict when something will be confirmed or not.