Name(s)

Krithi Koodli

Project Number

S1711

Project Title

A Novel Method of Distinguishing Rocky Planets from Uninhabitable Gas Giants Using Unsupervised Clustering

Abstract

Objectives

The objective is to develop a method of distinguishing rocky planets from gas giants. My engineering goals were to create a model using the Kepler data, confirm the model s validity by comparing its output to known planetary frequencies, and compare its predictions to 50 exoplanets that have been classified.

Methods

I used exoplanet data from the Kepler database. To classify exoplanets into the two categories, I needed to identify features in the data which could distinguish them. The main method of distinguishing between rocky and gas planets is to measure their densities. However, finding the density of a planet is difficult, as it requires finding the mass, which cannot be found from orbital equations. As a result, 900 exoplanets in the Kepler data have a measured density value, making their use in all 11,000 exoplanets infeasible. However, if I was to find rocky and gas planet clusters using density as a feature in the 900 planets, I could use them as a reference to evaluate clusters created by any feature for which data is available. Any feature that can distinguish between the two types of planets would have similar clusters to the density. I applied this to nine exoplanet features, and using the KMeans algorithm, I found that the most valid features were the planetary radius, stellar effective temperature, and orbital eccentricity.

Results

In order to determine the best feature, I applied the KMeans algorithm to the full Kepler data. I found that the frequencies of rocky planets and gas planets predicted by each feature were: 17% and 83%, 35% and 65%, and 38% and 62%, for the eccentricity, temperature, and radius, respectively. I then compared these to the frequencies observed by astronomers: 40% and 60%. Since the radius was closest to the actual frequency, I chose to use it to distinguish the exoplanets. I then tested the radius-based model on the 21 exoplanets that have thus far been confirmed to be rocky planets, and 29 exoplanets that have been confirmed to be gas planets. The model was able to accurately classify 100% of the rocky planets and 93% of the gas planets. This is a combined accuracy of 96% for the model s predictions, which thus validated the model. After completing my original plan, I found research which I am using to further validate my model.

Conclusions

The model met the engineering goals and could successfully distinguish rocky and gas planets. The radius is a strong feature for such a distinction, and it can be applied to exoplanetary data. I augmented the Kepler data with my model s predictions for each planet, which can be found at https://sites.google.com/view/exoplanetsorter-predictions.

Summary Statement

I designed a method of distinguishing rocky planets from gas giants by analyzing the Kepler exoplanet data.

Help Received

I researched and found the method myself, and my family and my mentor, Mr. Mostarshed, helped me review my results.