

Planet Hunter

Team Member: Richard Aguilar

Description

This project will involve the use of a machine learning to detect the existence of planets around other star systems. The hunt for extrasolar planets, or exoplanets, is a relatively new field of Astronomy. The first such planet was discovered in 1992. Since then, as of November 1st, 2020, over 4,300 exoplanets have been confirmed in over 3,200-star systems, with a little over 25% of those star systems being multi-planetary systems.

Detecting extrasolar planets is a difficult feat. These are objects that are quite literally outshone by their host star, and are relatively small in terms of size and mass. Thus, multiple methods have been developed to detect them. Three such examples are the radial velocity method, the microlensing method, and the transit method.

The radial velocity method hinges on Newton's laws of motion, specifically his third law: every action has an equal and opposite reaction. In this case, just as a star tugs on the planets orbiting it, so too does a planet tug on its star. The natural consequence of Newton's laws is that every pair of orbiting bodies in the universe revolve around a common center of mass: the barycenter. For bodies in which one is significantly more massive than the other, the common center of mass is physically inside the larger one. As such, what should happen is that the larger mass should appear to "wobble" as it revolves around the barycenter inside of it. This provides a useful way for detecting if a star is being orbited by a mass of significant magnitude—a planet.

The microlensing method is dependent on Einstein's General Theory of Relativity. His theory is a theory of gravity, providing an exclamation as to what "causes" gravity. Einstein posits that the gravity that we feel is not a "real" fundamental force, but, is in fact, a consequence of the geometry of spacetime. Any object with mass curves spacetime, distorting it. As such, when objects traveling in a straight line happen to stumble upon a sufficiently deep "well" caused by another body, if they are not traveling with sufficient velocity, they are caught within it, and travel in a geodesic (a straight line on a curved plane). This is what we interpret as gravity, and what we see as an "orbit". Since light is affected by this curvature, if an object with sufficient mass orbits a star, it should cause light emanating from the star to bend around it, distorting the image of the star. This also provides another useful way for detecting planets.

The third method, and the one this project will use, is called the transit method. This one is much simpler than the previous two and does not involve complicated physics. This takes advantages of the shadows cast on the host star from the smaller body. Every time a planet passes in front of a star, it causes a small dip in the apparent magnitude of the star. The amount of starlight blocked, and for how long it is blocked, is dependent on the size of the planet and its orbital period. If the star's light is blocked at regular intervals, it is likely that it is being orbited by a

planet. However, such blocking can also occur if it is being orbited by a smaller star (as is the case in a binary system), by asteroids and other bodies, as well as by gas clouds.

Member roles and contributions

This project will be a solo one, with myself as the researcher, programmer, and evaluator. I will implement the machine learning algorithm from scratch.

Enumerate CSCI 353 Topics

The algorithm in question will be the machine learning algorithm: Support Vector Machines (SVM). Since this is a binary classification task with data that is already labeled, using a supervised learning algorithm like SVM is appropriate. In addition, given the chance of data overlap, this was another reason to select it. SVM is a relatively flexible algorithm that can handle outliers and still provide a significant degree of accuracy. This is due to the use of Kernel Functions which elevate data to higher dimensions to more suitably classify them.

Describe dataset

The dataset is pulled from NASA's exoplanet archives. Specifically, it is coming from the Kepler Space Telescope, and contains 17 features, with 9,564 samples. This data was gathered over the course of 9.5 years from March 7, 2009 to October 30th, 2018, when Kepler was retired.

Describe timeline

Week of 12/01: Implement Support Vector Machine algorithm

Week 12/08: Fix any issues with implement, and write up report

Week of 12/15: Presentation

Describe demo

I will run the algorithm live, and show it classifying planets. I will also include graphs and perhaps other visualization tools to show the successes of the model.

Describe evaluation

Since this is a supervised learning project, evaluation will be relatively simple. The planets are already labeled into "CONFIRMED" and "FALSE POSITIVE". Therefore, I will evaluate the strength of the machine learning algorithm using the confusion matrix, and will see how high the F1 score is. I will also look at the individual combination of recall and precision.