CALIFORNIA SCIENCE & ENGINEERING FAIR
2019 PROJECT SUMMARY

Name(s)

David Freifeld

Project Number

J1407

Project Title

Detecting Exoplanets with Anomaly Detection and Hotelling's Theory

Abstract

Objectives

The objective is to determine which anomaly detection method is the most accurate at classifying lightcurves of stars, and which is the most suitable for use by astrophysicists.

Methods

A computer with Python 2.7 and multiple machine learning modules installed, and three base open-source machine learning programs were used. These were modified and tested for accuracy on labelled lightcurve data. I used cleaned lightcurve data from the Exoplanet Hunting in Deep Space challenge on Kaggle.

Results

Modified three programs to utilize anomaly detection strategies to classify lightcurves, and were tested on a labelled dataset. Many tests were run with different input parameters to determine the best performance of each strategy. Hotelling's Theory proved to be the best, having the highest accuracy of 74.7% when used with a threshold of 0.3, a dip threshold of 175, and a cutoff of 1500.

Conclusions

Hotelling's Theory ran significantly faster than the other two programs and was the most accurate showing that it was a effective method for detecting exoplanets. It's speed shows that it is suitable for use on large amounts of data, and these two features allow a significantly higher amount of exoplanets to be discovered in a shorter timeframe than the traditional manual method.

Summary Statement

I found that Hotelling's Theory is a suitable method for detecting exoplanets automatically as shown by high testing accuracies and fast runtime.

Help Received

Andrew Lookingbill advised me on how to set up LSTMs and RNNs, along with Conner Vercellino who advised me to begin the project focusing on the dataset.