Using Machine Learning to Identify Young Stars with TESS Data

C. Santiago[1], A. Chaushev[2], S. Sallum[3]
[1]UC-Irvine

Abstract

While most studies use stellar colors to find young stars, we measure the variability characteristics of known young stars, and compare them to other stars using an algorithm we constructed to group similar objects together based on their lightcurves. In the algorithm we use the Lomb-Scargle periodogram to convert the data into the frequency domain. We then apply machine learning techniques from the scikit-learn library - Principal Component Analysis (PCA) and T-distributed Stochastic Neighbor Embedding (TSNE) - to create a metric to help identify groups of similar stars. We make use of two multi-sector datasets: the Cluster Difference Imaging Photometric Survey (CDIPS) and the Transiting Exoplanet Survey Satellite (TESS) main survey. With the help of UCI's High Performance Community Computing Cluster (HPC), we present a preliminary analysis of several sectors of data from TESS.

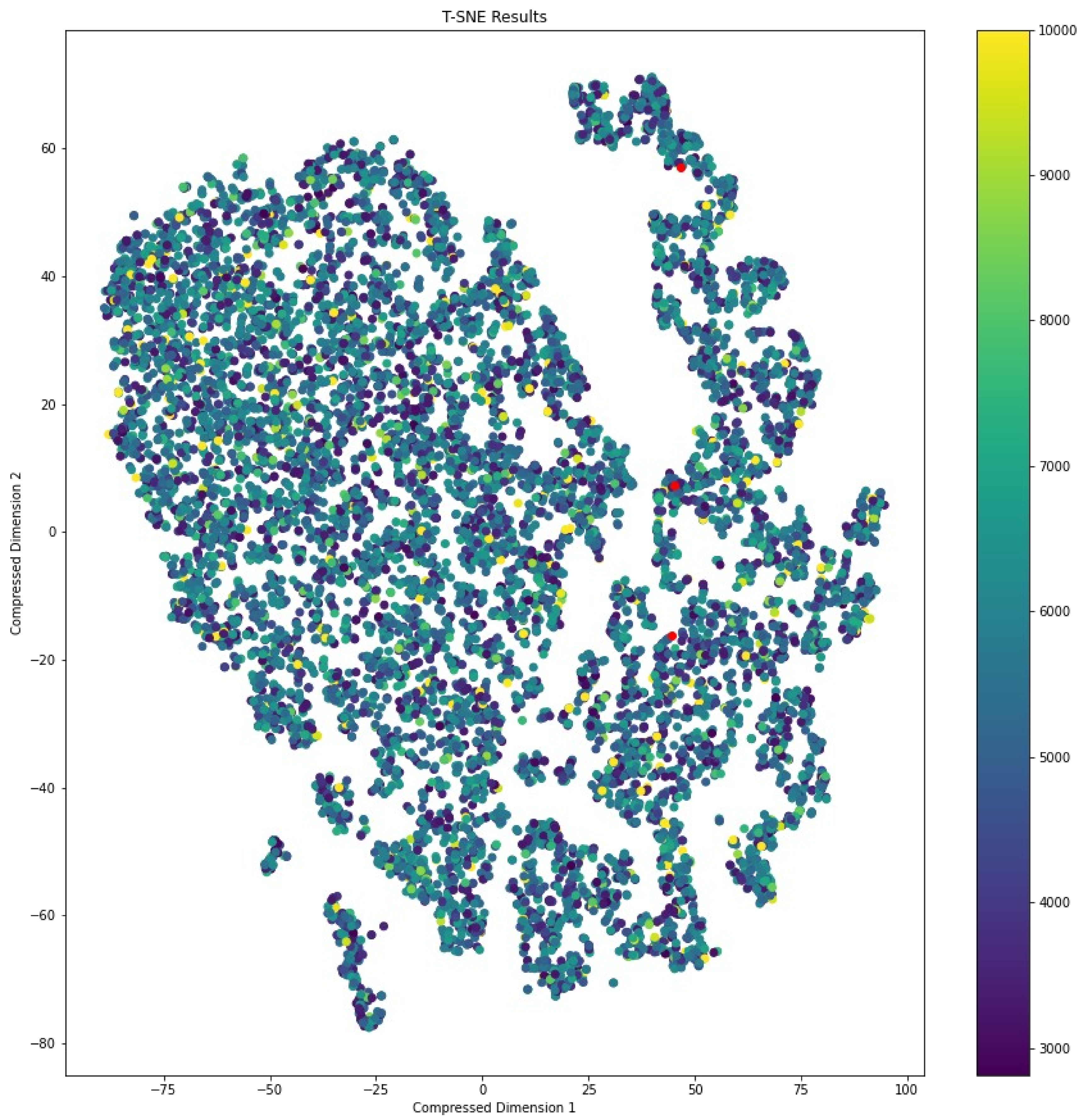
Motivation

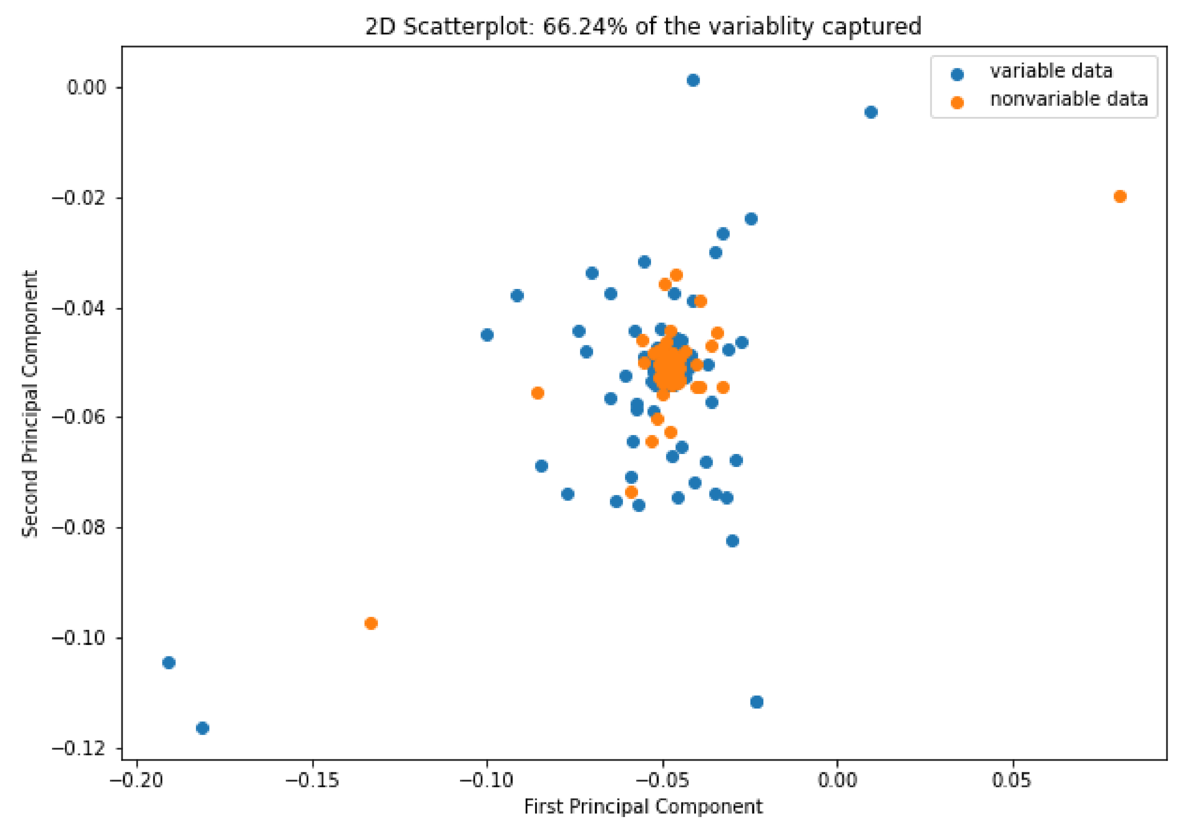
Observations of young stars are essential to developing our understanding of planet formation. Many young stars are found together in well-studied groups, clusters and associations. We aim to identify new candidate young stars, not necessarily in well-studied associations, which may otherwise be missed.

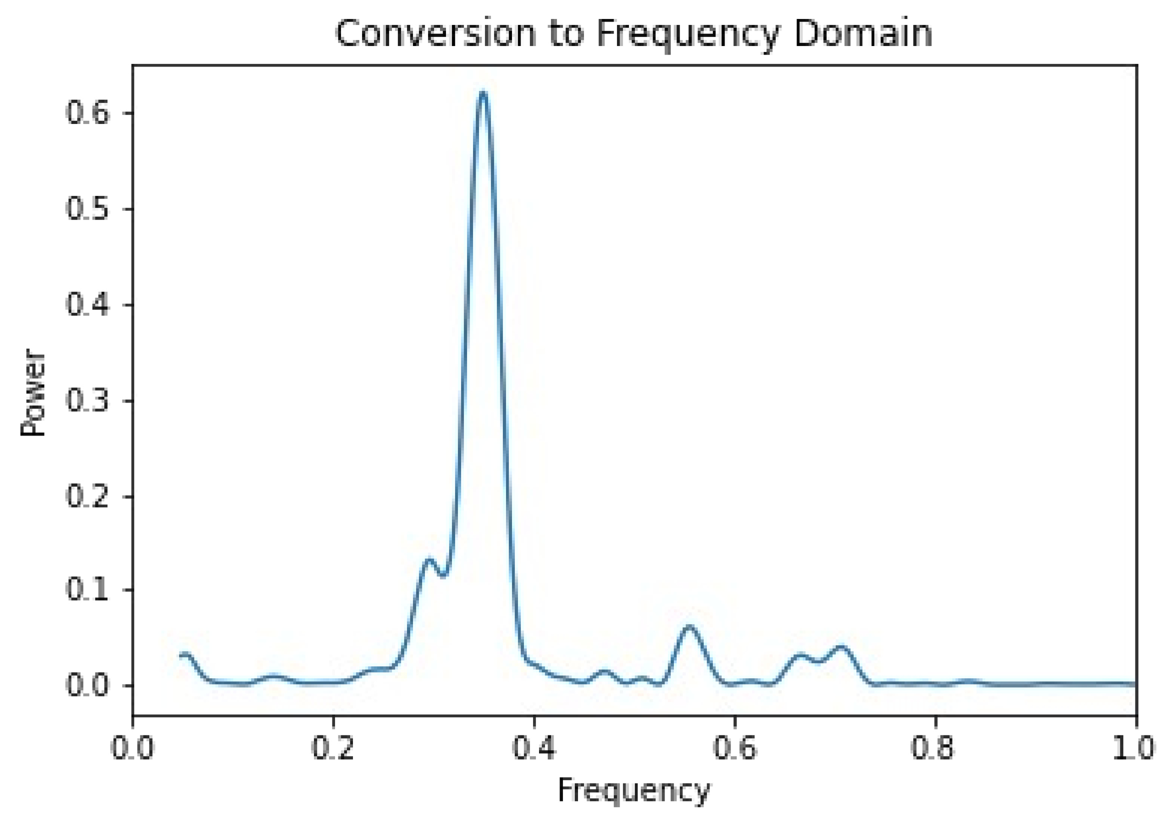
Methods

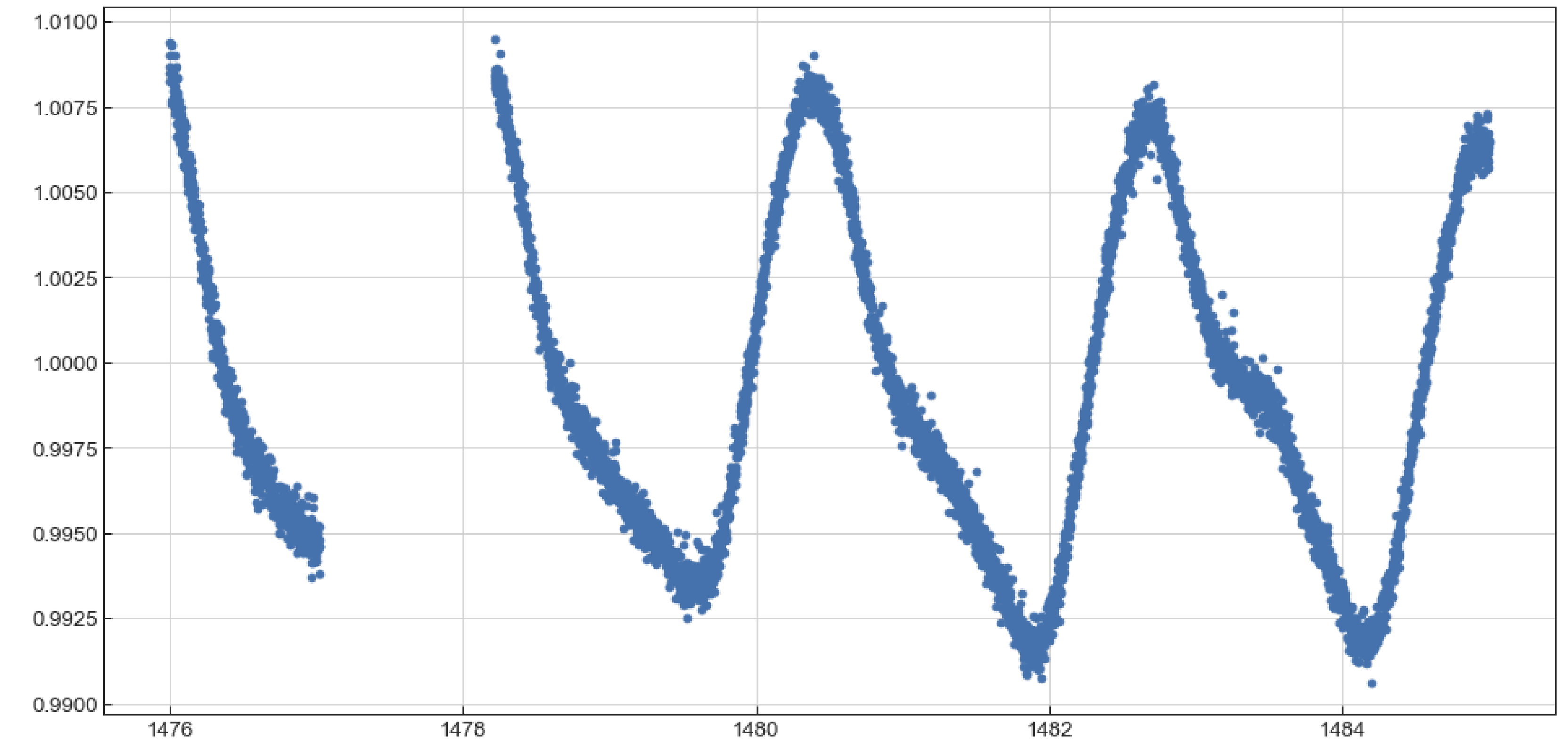
We focus on the variability of the light curves for our findings since stars with similar variability are likely to have the similar age. With this we choose to back process our data to compensate for our limited memory capacity. The data we processed were light curves from TESS sector 6, and primarily processed the flux and time of these light curves. Afterwards we shifted our data from the time domain to the frequency domain, the purpose of this was having a more accurate result once our data has been through dimensionality reduction through the PCA. We choose to have our data reduced to 10 components since our data is best expressed to 10 principal components. Having the PCA fitted to 100 variable and nonvariable light curves. After all our data has been through the PCA we then used TSNE to further reduce the dimensionality of our data to two components to better visualize.

Results









Conclusions

From this we could further analyze the rest of the TESS sectors to further develop our metric and optimize it for better results.

References

Tofflemire, Benjamin M., et al. 2021, "TESS Hunt for Young and Maturing Exoplanets (THYME). V. A Sub-Neptune Transiting a Young Star in a Newly Discovered 250 Myr Association". The Astronomical Journal, Volume 161, Issue 4, id.171, 32 pp.
Doi: https://ui.adsabs.harvard.edu/link_gateway/2021AJ....161..171T/doi:10.3847/1538-3881/abdf53

Ricker, George R. et al. 2015, "Transiting Exoplanet Survey Satellite (TESS)". Journal of Astronomical Telescopes, Instruments, and Systems, Volume 1, id. 014003
Doi: https://ui.adsabs.harvard.edu/link_gateway/2015JATIS...1a4003R/doi:10.1117/1.JATIS.1.1.014003