PREDICTING POSSIBLE EXOPLANET HOSTS BY APPLYING MACHINE LEARNING TECHNIQUES TO STELLAR ABUNDANCES . N. R. Hinkel[1] and R. Galvez[2], [1]Department of Physics & Astronomy, Vanderbilt University, Nashville, TN 37235 (natalie.hinkel@gmail.com), [2]Department of Physics, Fisk University, Nashville, TN 37208.

Introduction: Since the initial discovery of the "planet-metallicity" relationship for giant planets, as popularized by Fischer & Valenti [1] the iron content within a star, or the [Fe/H] ratio, has been used a proxy for the overall metallicity of the star. While it has been assumed that the abundances of other important bio-essential elements, such as C, O, Si, and Mg, have been consistent with the Fe-trends in giant planet hosting stars, these results have not been seen despite a variety of studies over the last ~10 years. Additionally, there has not been any detected correlation between stellar abundances and smaller, terrestrial planets.

Despite the huge number of exoplanetary detections from the Kepler mission, the traditional radial velocity and transiting detection techniques only utilize the physical properties of the stellar system. Here we take advantage of the *host's stellar abundances* in order to statistically examine any possible dependence of the occurrence of exoplanets to the chemistry of the star. We used the Hypatia Catalog [2, 3] as a large sample of non-Fe abundances for stars that do and do not have detected exoplanets. We produced a target list of possible planet-hosting stars that have a high probability of hosting a detectable exoplanet.

Hypatia Catalog: The Hypatia Catalog is a database of stellar abundances which includes +65 elements and species within >6000 FGK-stars that are less than 150 pc from the Sun. Hypatia was compiled from over 200 literature sources such that the data were homogenized to the same solar scale. The median value was used during those instances where multiple literature values existed for the same element in the same star. Hypatia currently contains stellar abundances for +300 exoplanet host stars.

Supervised Machine Learning: When comparing non-Fe element ratios in stars with and without planets, it is clear in our analysis that, on average, detected exoplanet hosts have higher abundance values. We have leveraged this observation to train a supervised algorithm as a binary classification problem, where the solar normalized [X/H] abundance data are used as predictive features for a likely planetary detection. The algorithm is comprised of multiple "weak" classifiers (such that the error rate is only slightly better than random) in order to form a powerful "committee" of algorithms, also known as a "boosted" family of classifiers.

The algorithm collects a random sub-sample of detected exoplanet hosts in addition to background stars (or stars without a detected planet) in equal numbers, in this case, 300 stars. The Exoplanet Orbit Database [4] was used when determining whether a star had a planet. A classifier is then trained on a sample of data where the prediction (whether or not the star has a planet) is known. This process is done for 1000 iterations, assessing the performance of our predictor, and calculating the mean probabilities until our model scores no longer changed per iteration. In this way, we were able to produce a metric that determines the probability of an existing exoplanet while averaging out noise inherent to the data.

Conclusion: Our model had a very striking "positive rate" of ~85%, or when the model was able to correctly predict the existence of an exoplanet. In addition, the algorithm was able to correctly predict a "true negative," or when a star with no observed planet is predicted to not have a planet, ~66% of the time. Note that, due to the limits in planetary detectability, we are unable to adequately determine a true-negative since the presence of a planet cannot currently be excluded. The success that we have achieved with our model is likely due to the large range in non-Fe element abundances. In addition, while we have confirmed that Fe is important to predicting exoplanet hosts, a surprising result has been that Na is also statistically influential. We will discuss this result, in addition to our list of likely exoplanet hosts stars.

References: [1] Fischer, D. and Valenti, J. (2005), *ApJ, 622*, 1102. [2] Hinkel, N. R., Timmes, F. X., Young, P. A., Pagano, M. D., & Turnbull, M. C. (2014), AJ, 148, 54. [3] Hinkel, N. R., Young, P. A., Pagano, M. D., et al. (2016), *ApJS, 226*, 4. [4] Han, E., Wang. S.X., Wright, J.T. et al. (2014) *PASP, 126*, 943.