A Machine Learning Approach to Determine the Number of Exoplanets in a Given Planetary System Based on NASA Exoplanet Archive Data

Zhang, Maureen (School: Pittsford Sutherland High School)

Accurately determining the number of exoplanets in planetary systems is essential for gaining insight into planetary formation, the search for potentially habitable exoplanets, and the design of future exoplanet missions. Using data on exoplanets from the publicly accessible NASA Exoplanet Archive, unsupervised and supervised machine learning methods were applied to determine the number of exoplanets in a given planetary system. Clustering methods were applied in order to understand the multidimensional dataset and evaluate the potential efficacy of machine learning methods. The accuracies of models applied to this multiclass classification problem with imbalanced classes were obtained and reported. Applying algorithms including a decision tree, logistic regression, random forest, and support vector machine resulted in model accuracies comparable to similar studies where machine learning was applied to exoplanet data (Schanche et al., 2019). A weighted average ensemble was then applied to the models in order to achieve better classification accuracy. While the resulting model accuracies are on par with previous studies of exoplanet data (Schanche et al., 2019), higher accuracies may be achieved with the incorporation of additional data from the James Webb Space Telescope (JWST) to this project. Although these models were successful, due to the ever-changing nature of data on the number of planets present in planetary systems, cluster labels returned from clustering on the data were used as a target variable as a second approach. A significant increase in accuracy resulted, indicating that the clustering methods used hold great potential in identifying undetected exoplanets within known planetary systems and could be used to inform future exoplanet missions.

Awards Won:
University of Texas at Arlington College of Science: Alternates in case any of your recipients do not accept the award (not to be read aloud)
Second Award of $2,000