catanzarite, Joseph
Automatic Classification of Kepler Planet Candidates

We have implemented a machine learning technique to produce a catalog of planet candidates from the Kepler Q1-Q17 TCEs (Threshold Crossing Events). Classification is accomplished by means of a machine learning technique known as the random forest. The inputs to our classifier (called the autovetter) are a training data set composed of TCEs that have been dispositioned (mostly) by humans, and a set of attributes associated with each TCE. From the training set, the autovetter 'learns' a mapping between attributes and predicted classification. This mapping is then applied uniformly and consistently to all TCEs to produce a catalog of planet candidates. The autovetter catalog is distinct from the robovetter catalog, the Q1-Q17 DR24 KOI (Kepler Object of Interest) table provided at NExScI (NASA Exoplanet Science Institute). The autovetter and robovetter have followed different approaches to arrive at the same goal – automation of the process of human classification of planet candidates to achieve fast, robust and consistent vetting of the entire population of TCEs. While the robovetter's heuristics are 'hardwired' by humans, the autovetter's are 'learned' autonomously from the data. A major difference between the catalogs is that the autovetter also provides a Bayesian measure of the confidence (posterior probability) that each object is a planet candidate. Posterior probabilities can be advantageously used in statistical studies such as occurrence rate calculations to de-weight planet candidates that are at the noisy edges of the planet catalog.