Machine Learning: Findings and Predictions

1. Decision Tree
   1. 87% accuracy after tuning hyper parameters with GridSearchCV
2. Random Forest
   1. 87% accuracy after tuning hyper parameters with GridSearchCV
3. K Nearest Neighbors
   1. 82.5% accuracy after tuning hyper parameters with GridSearchCV
4. Neural Network & Deep Learning
   1. 87% accurate after incorporating additional deep learning filters

Decision Trees, Random Forests, and Neural Networks can accurately find patterns in planet data to determine habitable candidacy with greater than 85% accuracy. We can use machine learning to run through data to narrow down search items and further reduce workload. K Nearest Neighbors seems to be less effective, suggesting that the data can’t be accurately expressed using KNN.

Based on the Neural Network model, the differentiation between ‘CANDIDATE’ and ‘CONFIRMED’ is closer than ‘FALSE POSITIVE’

This suggests there’s a way to try and train the deep learning model based solely on CANDIDATE vs. CONFIRMED data as a way to improve the model.

Top categories considered as influential based on both the Decision Tree and Random Forest includes:

Decision Tree:

1. Koi\_fpflag\_ss
2. Koi\_fpflag\_co
3. Koi\_fpflag\_nt
4. Koi\_model\_snr
5. Koi\_fpflag\_ec

Random Forest:

1. Koi\_fpflag\_ss
2. Koi\_fpflag\_ss
3. Koi\_prad
4. Koi\_fpflag\_nt
5. Koi\_model\_snr