CLASSIFYING TESS LIGHT CURVES OF VARIABLE STARS WITH MACHINE LEARNING

Alex Parsells, Dept. of Physics and Astronomy, University of Oklahoma, Norman, OK 73019
Mentors: Daniel Giles and Ann Marie Cody, SETI Institute, Mountain View, CA 94043;
Steve Croft, Berkeley SETI Research Center, Berkeley, CA 94720

Introduction

- The Transiting Exoplanet Survey Satellite (TESS) launched in 2018 and produces optical light curves (brightness over time) across almost the entire sky
- Here I use machine learning to classify TESS light curves into different classes of variable stars (e.g., eclipsing binary, pulsator, etc.)
- Classifications will be used for a complementary anomaly detection project, which will search for light curves that may be either members of rare or unknown variable types or candidates for transiting megastructures (SETI interest)
- Goal: classify 3+ variable types, each with 90%+ accuracy, on sectors 1-26 of TESS data

Methods

- First variable type: eclipsing binaries (EBs) from Villanova *Kepler* catalog (2922 objects)
- Data cleaning: converted to TESS IDs, restricted to TESS sector 14 (for constant systematic effects) and kept only those between TESS magnitude 10 and 15. Left with 2172 objects
- Summarized light curves into 29 numerical features with Cesium
- Machine learning with scikit-learn: support vector, random forest, and k-nearest neighbors classifiers on EB vs. non-EB. Grid sweeps over hyperparameters with Weights & Biases

Results

- Maximum accuracy around 80% with each of three classifiers
- Precision typically between 0.7-0.8, recall up to approximately 0.9

Discussion/Next Steps

- Current performance leaves room for improvement
- Filter out known long-period EBs from data (may not be detectable as EBs from TESS) -- this should improve accuracy, hopefully past 0.8
- Try other premade and custom features on light curves (e.g., m-statistic)
- Classify other variable star types (next: flare stars)
- Strategic Minority Oversampling Technique (SMOTE) for fixing imbalanced data
- Experiment with other classifiers (including neural networks)

References

- Biewald, L. 2020, *wandb.com*
- Debosscher, J., Blomme, J., Aerts, C. & De Ridder, J. 2011, *A&A*, 529, A89
- Giles, D. & Walkowicz, L. 2020, *MNRAS*, 499, 1
- Kirk, B., Conroy, K., Prša, A., et al. 2016, *AJ*, 151, 3
- Lightkurve Collaboration. 2018, *Astrophysics Source Code Library*, record ascl:1812.013
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12
- Yu, L., Vanderburg, A., Huang, C., et al. 2019, *AJ*, 158, 1

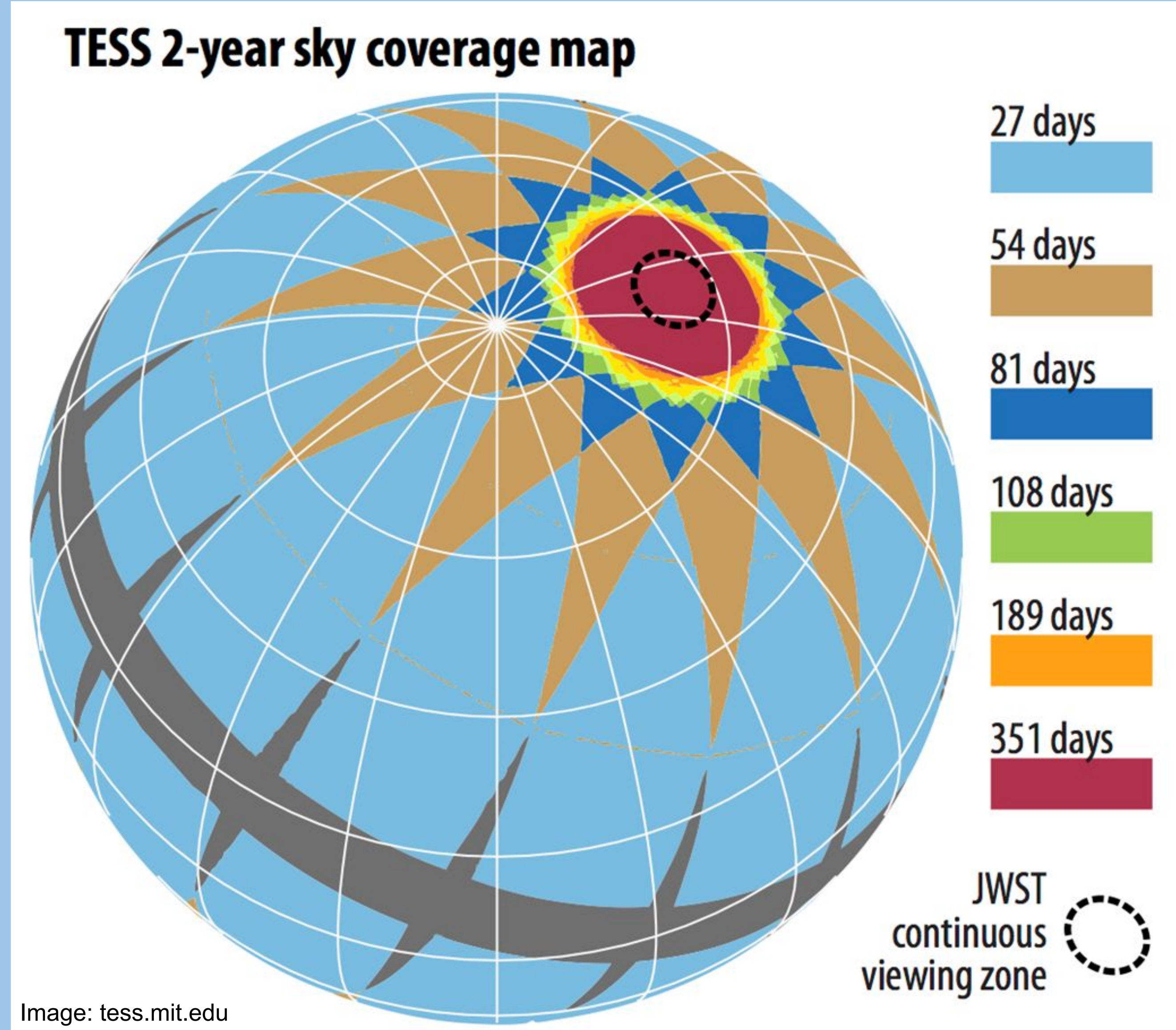
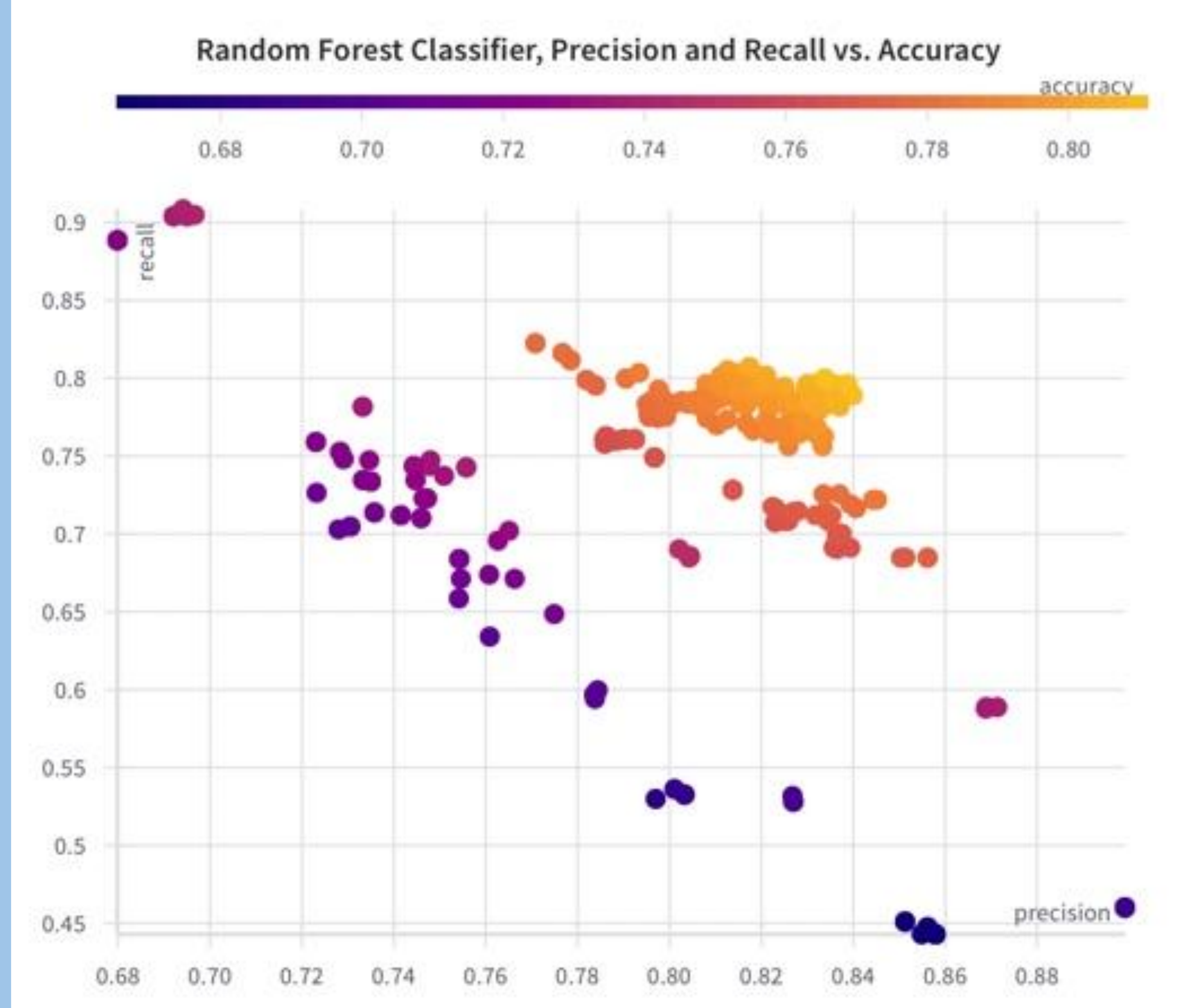
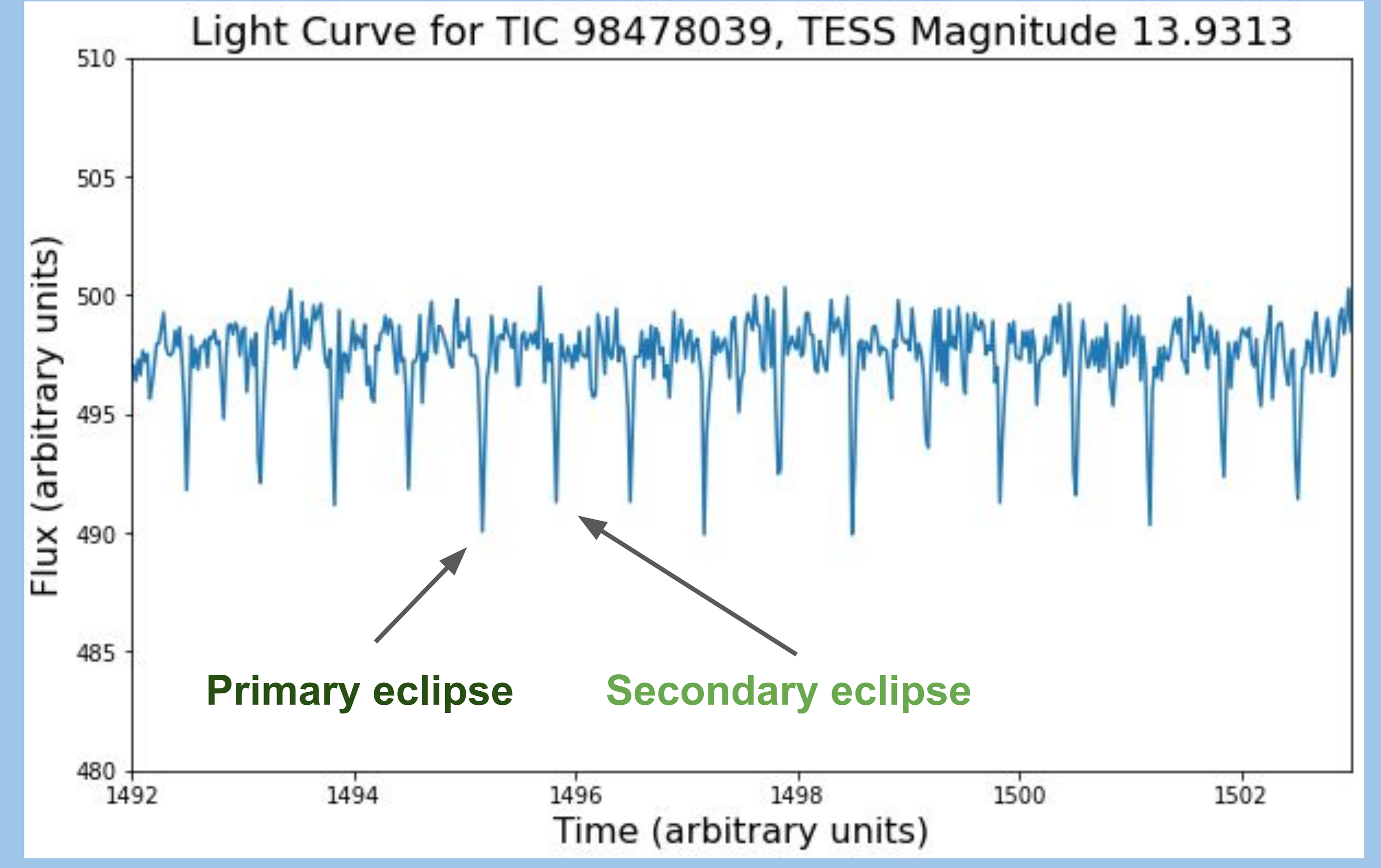
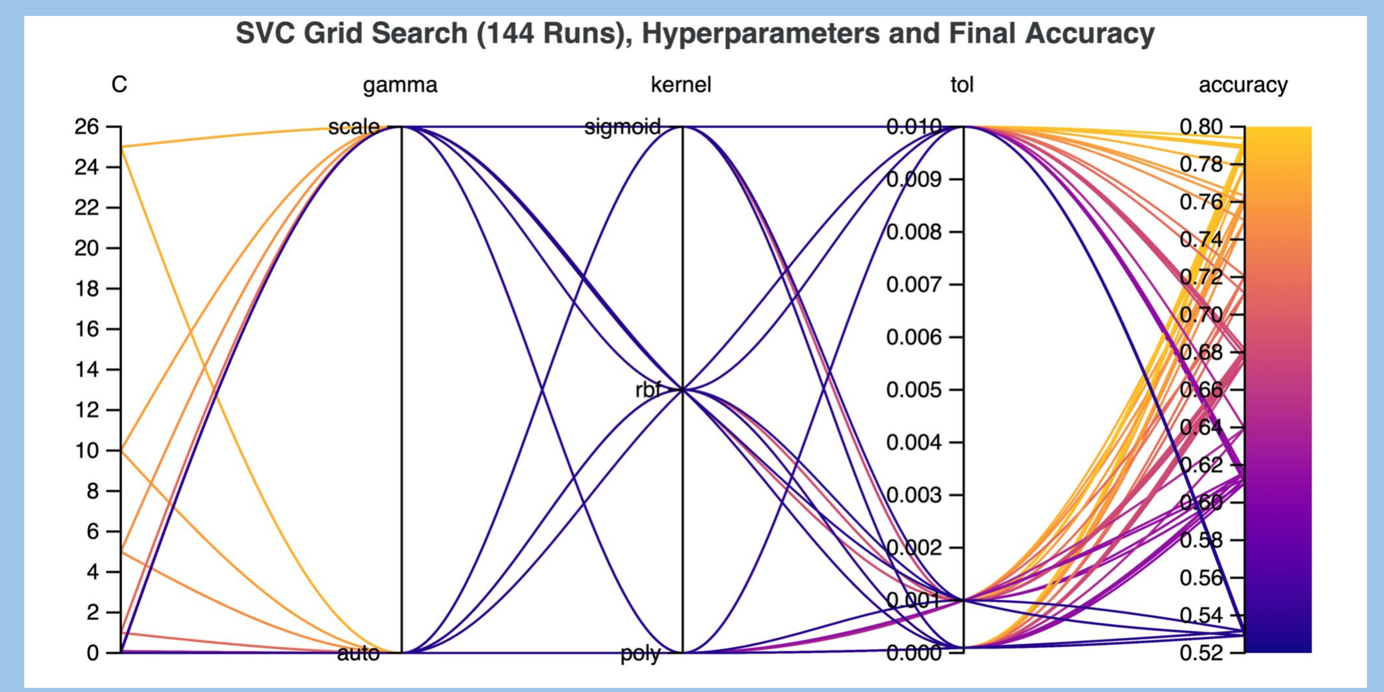
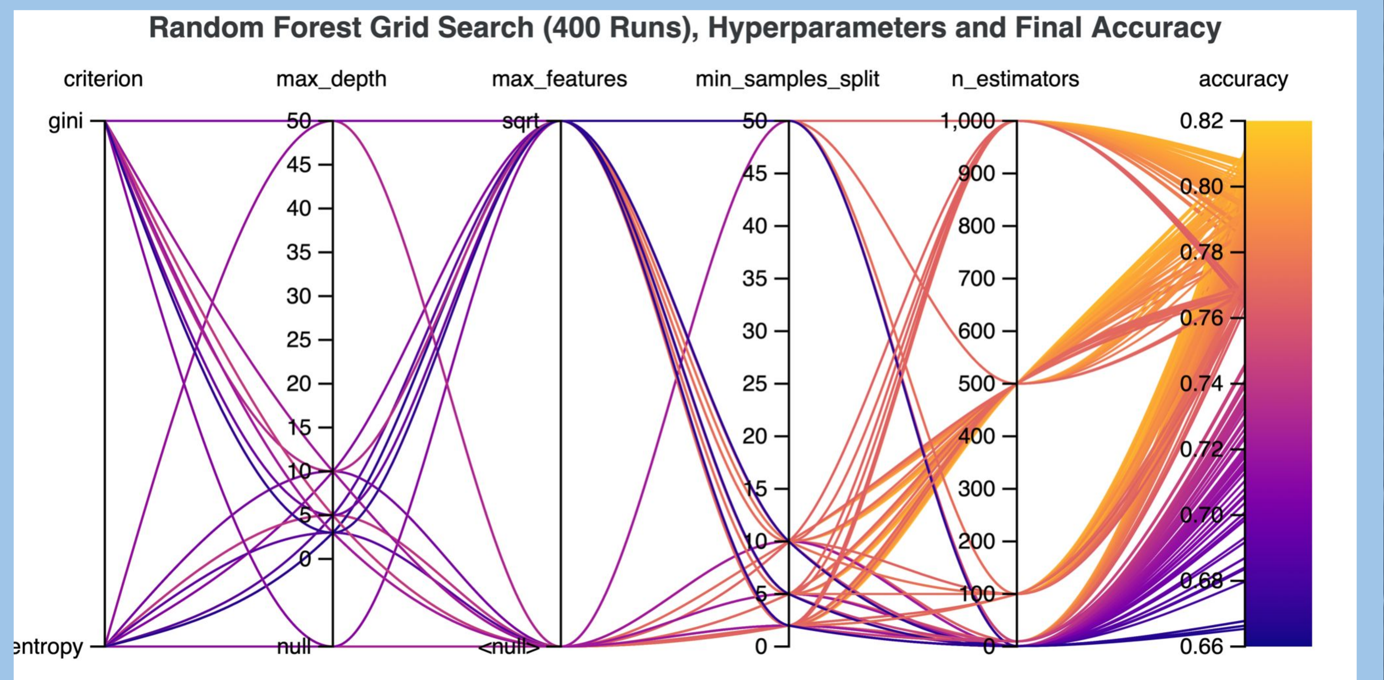


Image: tess.mit.edu









Acknowledgments

I would like to thank my mentors: Daniel Giles, Ann Marie Cody, and Steve Croft, for introducing me to TESS light curves and machine learning, and for guiding me and providing encouragement throughout the project. I thank Brian Powell and Ethan Kruse for making available their classifications of TESS eclipsing binaries for use in creating a test dataset. I also thank the various faculty, staff, and current and former interns affiliated with the Breakthrough Listen program who helped me troubleshoot technical issues or suggested better solutions to problems along the way. These people include Jim Davenport, Dave MacMahon, John Hoang, and Howard Isaacson. Finally, I gratefully acknowledge financial support from the National Science Foundation through its REU grant to Berkeley SETI Research Center.