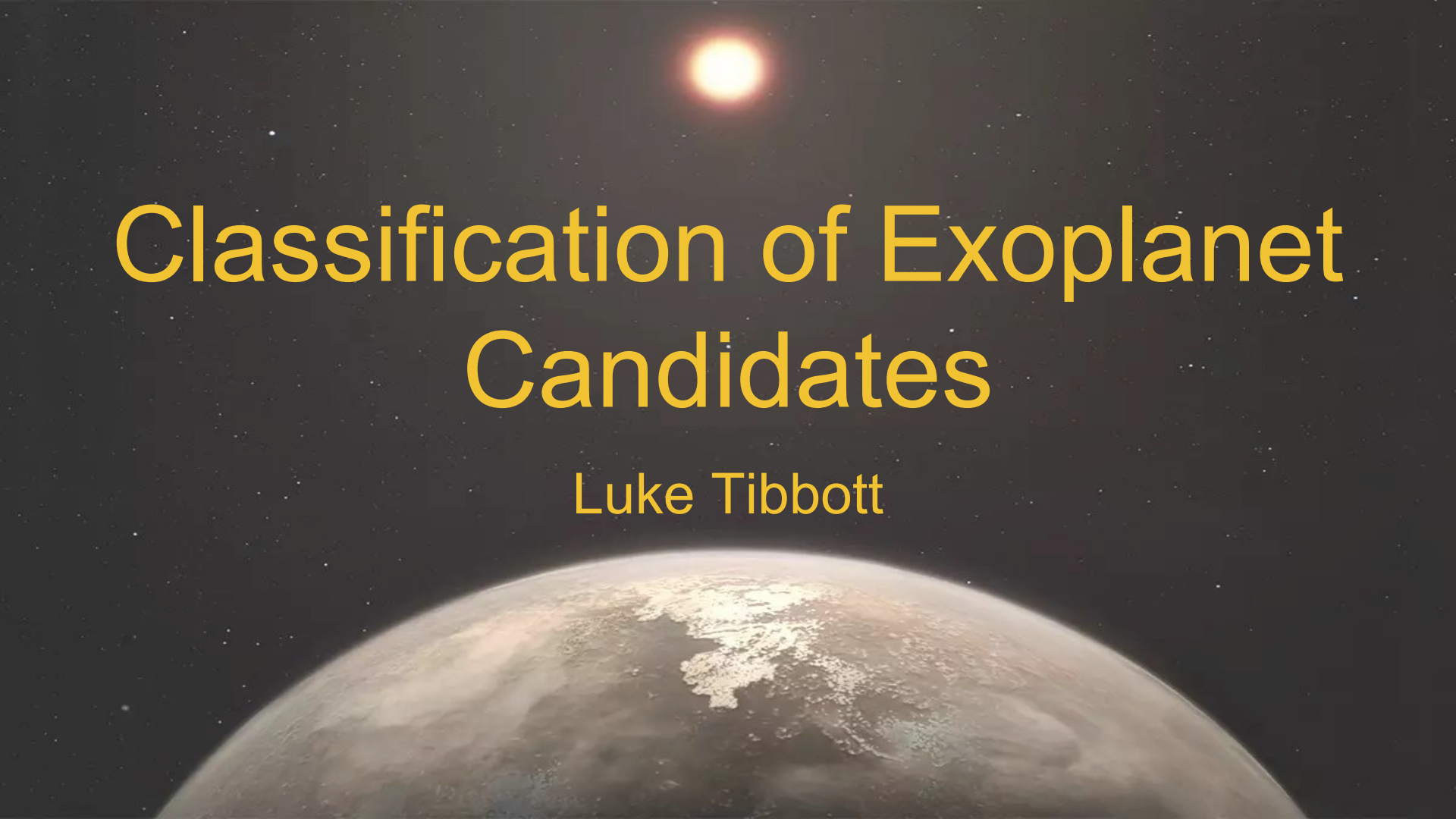
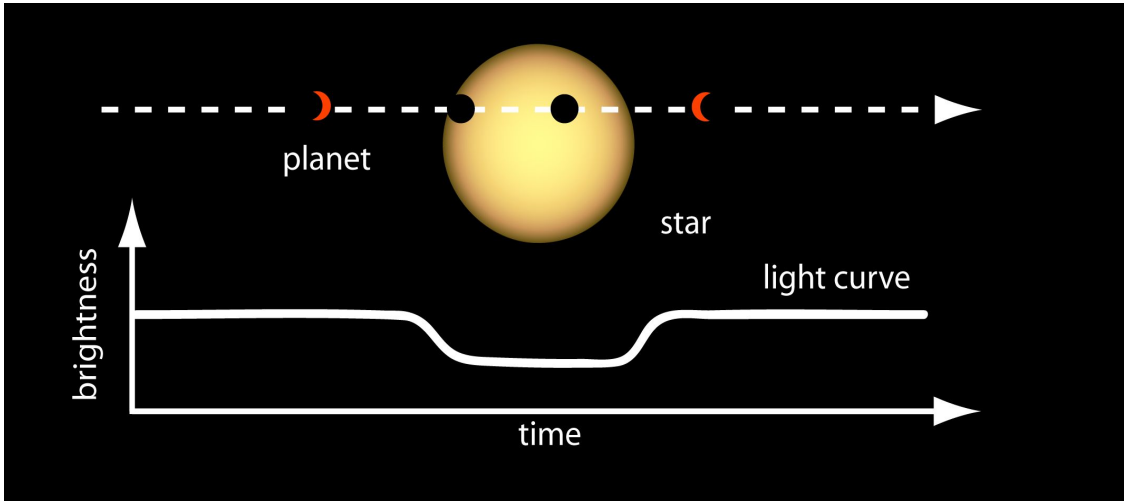


Classification of Exoplanet Candidates

Luke Tibbott

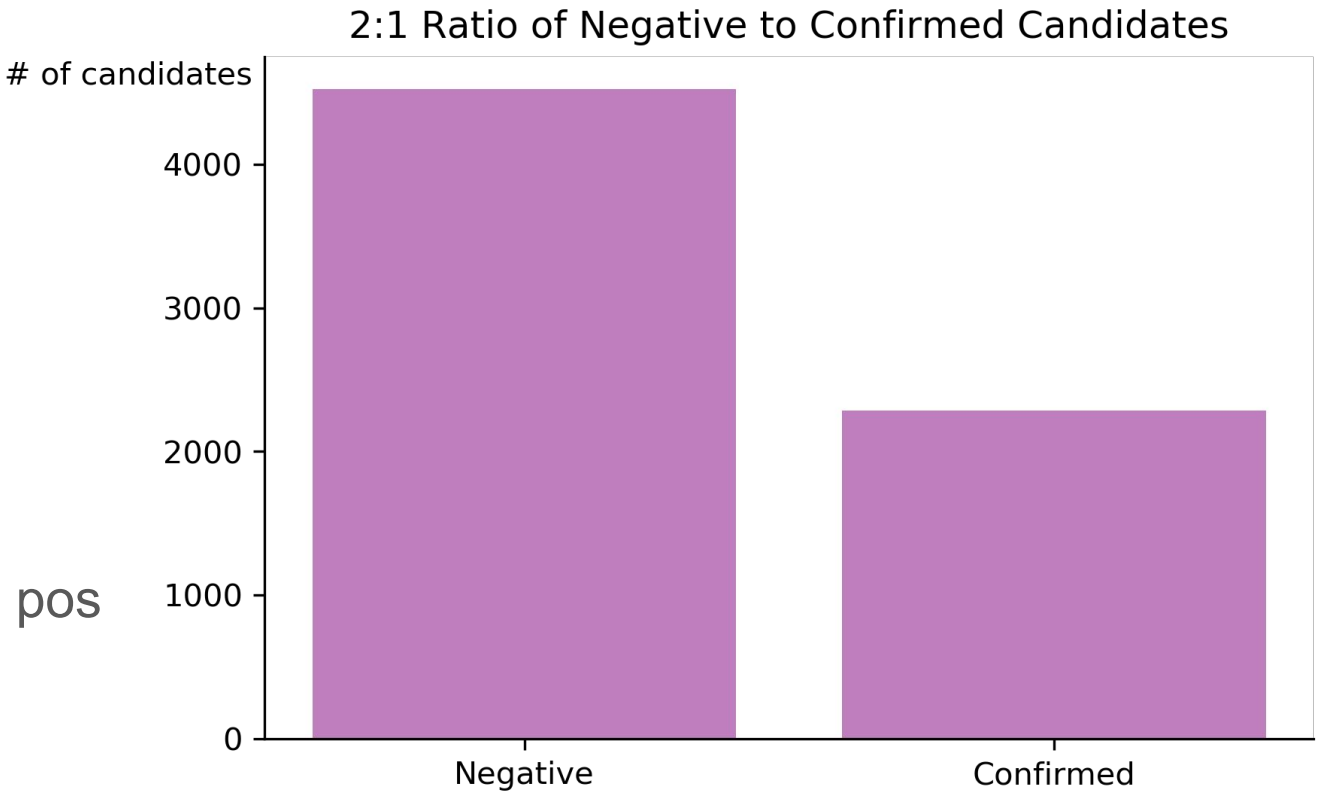
Classification the ol’ fashioned way

- Transit method
- Light curve analysis
- Machine learning?



Data

- NASA
- 7,000 candidates
- 2 : 1 ratio of neg : pos

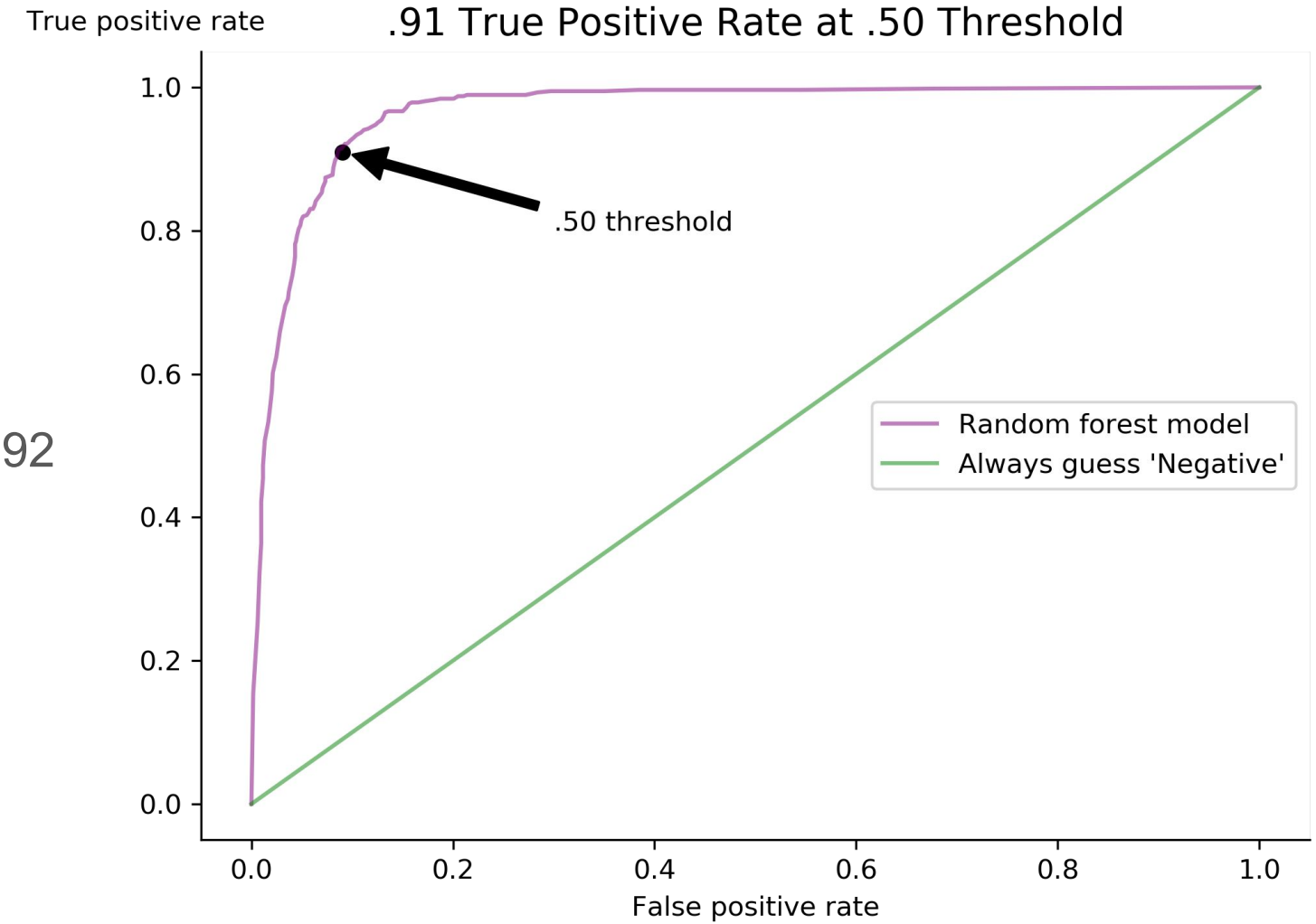


Modeling

- Random Forest Classifier
- Minority Oversampling
- Gini importance

Results

- Recall - .91
- Precision - .92
- AUC - .90



Conclusions

- Machine learning very effective
- Cheap alternative to light curve analysis
- Predictions available through Flask app

Predict an exoplanet candidate's disposition!

Planet period: 3 days

Impact parameter: .5

Duration of transit: 5 days

Stellar optical depth: 700

Planet radius: 4 Earth radii

Planet equilibrium temperature: 500 Kelvin

Insolation flux (as a ratio to that of Earth-Sun system): 8

Model signal to noise ratio: 30

Get prediction

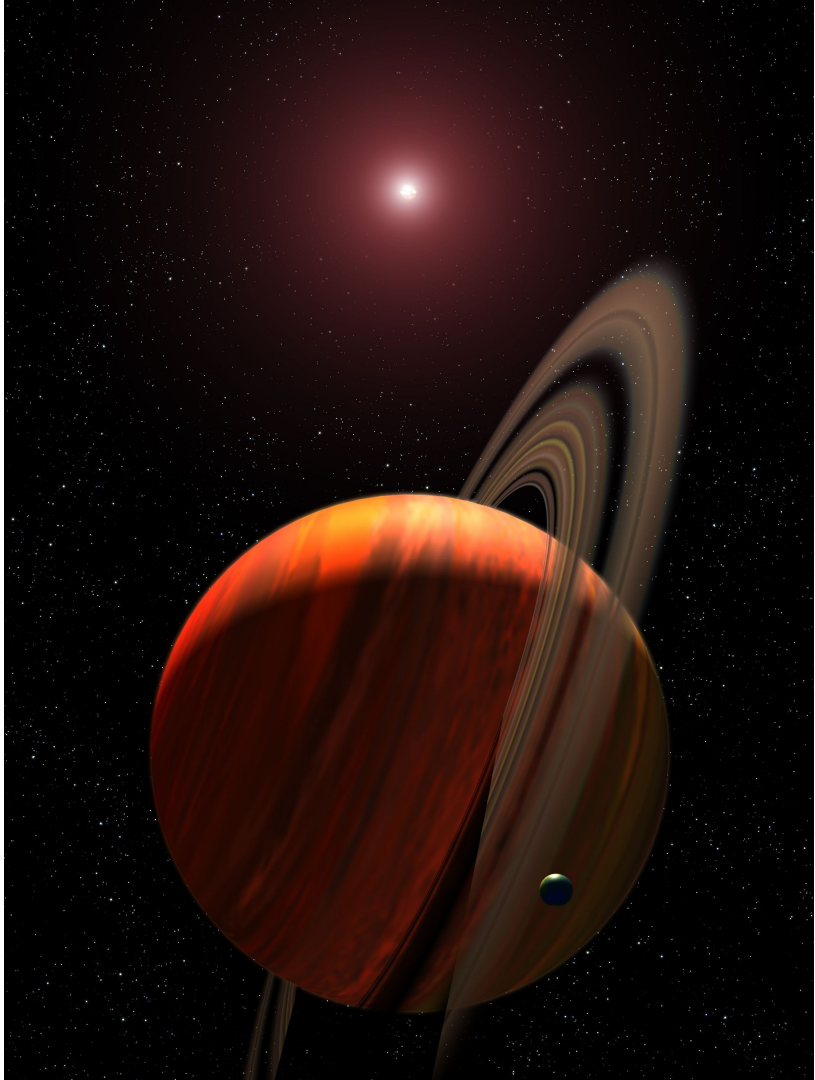
Our prediction is:

This candidate is **an exoplanet**

Probability of candidate being an exoplanet: **0.85**

Future work

- Neural networks
- Improve domain expertise
- Very active problem area



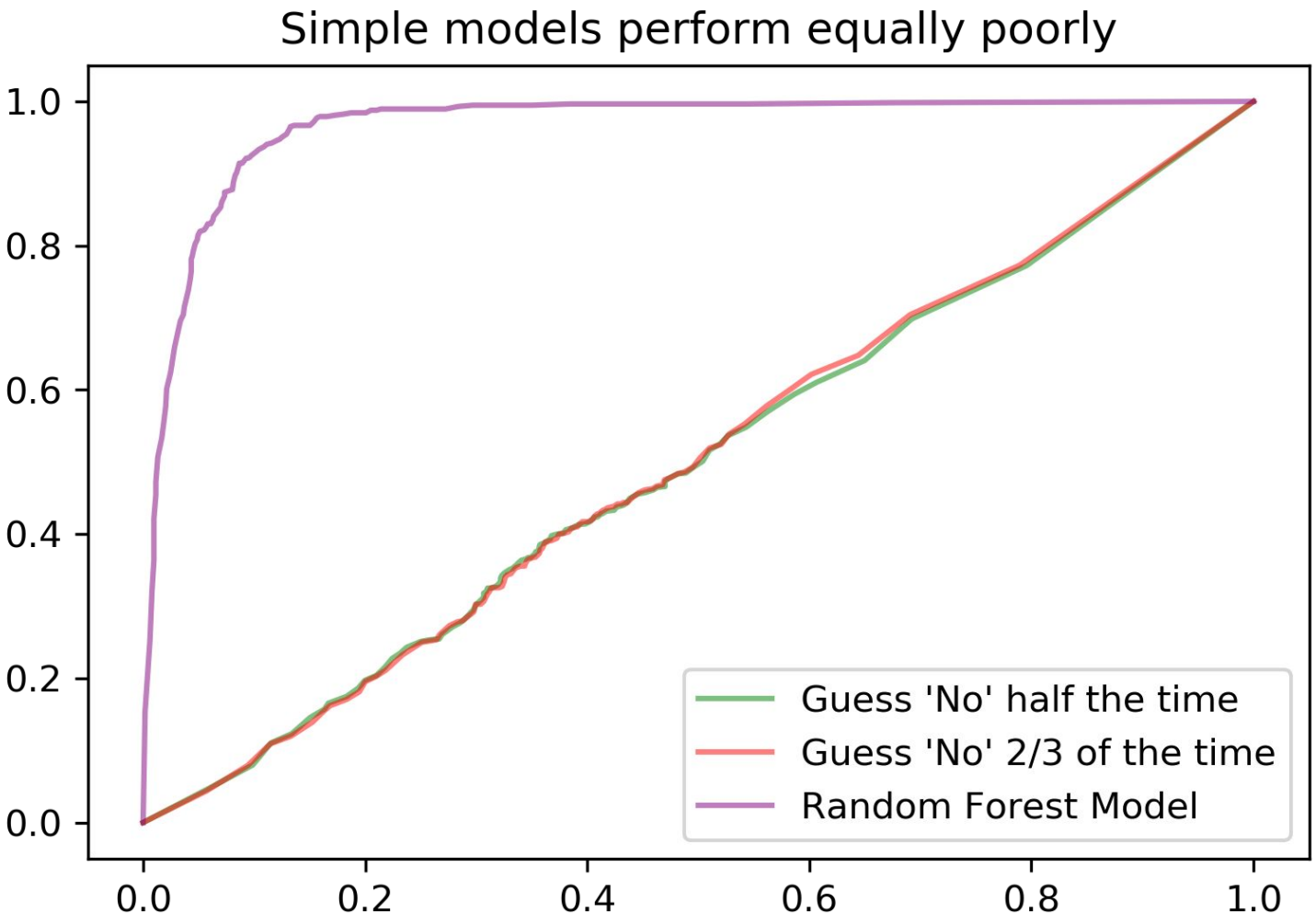
Thank you!

luke.tibbott@gmail.com

github.com/luketibbott

https://www.linkedin.com/in/luke-tibbott

Appendix

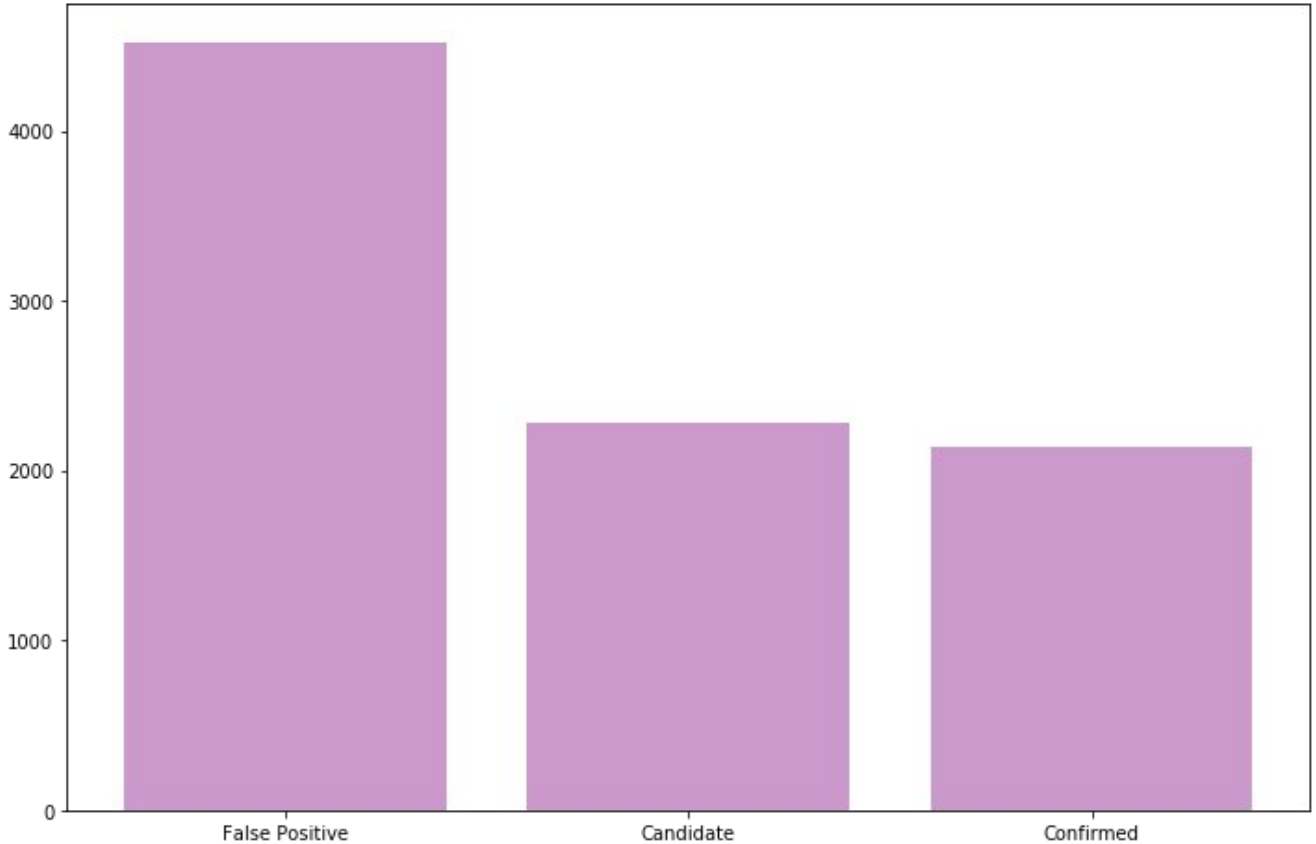
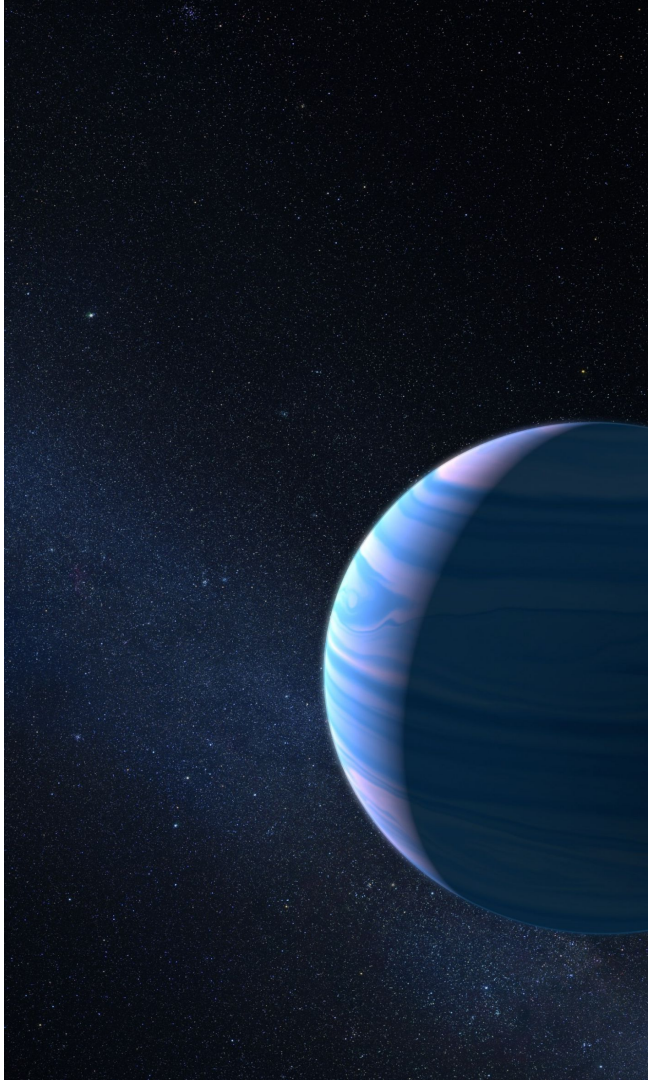

Simple models perform equally poorly
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Guess 'No' half the time
Guess 'No' 2/3 of the time
Random Forest Model

Classification report

```
Features sorted by their score:
[(0.2149, 'prad'), (0.1939, 'model_snr'), (0.1292, 'impact'), (0.1121, 'period'), (0.0954, 'depth'), (0.0893, 'inso
l'), (0.0873, 'teq'), (0.078, 'duration')]
[1 1 0 ... 0 1 0]
              precision    recall  f1-score   support

           0       0.95      0.91      0.93      1131
           1       0.84      0.91      0.88       572

   micro avg       0.91      0.91      0.91      1703
   macro avg       0.90      0.91      0.91      1703
weighted avg       0.92      0.91      0.91      1703

Best parameters:  {'randomforestclassifier__n_estimators': 100, 'randomforestclassifier__max_features': 2, 'randomfor
estclassifier__max_depth': 100}
Mean ROC grid scores:  [0.9372237  0.93854996 0.93825523 0.94002358 0.93766578 0.93884468
 0.9360448  0.93265547 0.93692897 0.93869732]
```

Disposition counts with candidates included