Machine Learning for Physics and Astronomy: Exercises

David Sergio

July 24, 2025

Problem 3.4

Based on your knowledge of the laws of physics (in particular, the scaling of stellar luminosity as a function of mass for main sequence stars and the scaling of flux from a point source with distance), engineer a new feature (a combination of two or more features you already have) that might correlate strongly with planet habitability. Explain why you think this feature would work. [1]

– extra points if you include a special reference to Battlestar Galactica and their ability to live on a planet but not one like earth. . . . that might not even make sense but something like that lol

Background

The mass-luminosity relationship for main sequence stars is given by the equation:

$$L \propto M^{3.5} \tag{1}$$

where L (watts) is the luminosity and M (kg) is the mass of the star. [4]

The flux F from a point source at a distance d is given by:

$$F = \frac{L}{4\pi d^2} \tag{2}$$

where F (watts per square meter) is the flux and d (meters) is the distance from the source. [5]

Tylium and Battlestar Galactica

In the universe of Battlestar Galactica, Tylium is a valuable resource used to fuel spacecraft. [2] A star system with a high concentration of Tylium will support advanced technology and space travel, critical for the survival of humanity and winning the Cylon War. If we assume Tylium is a metal (a shorthand for non-hydrogen and non-helium), it follows that planets in star systems with high metallicity would have a higher concentration of Tylium, either on the planet or on a nearby moon, asteroid, or neighboring planet. Taking this into account, we will need to limit our search to stars with metallicity > 0.00, where metallicity is defined logarithmically, and scaled relative to solar metallicity [3], as these stars are more likely to have planets with Tylium deposits.

$$\frac{Fe}{H} = \log_{10}\left(\frac{(Fe/H)_*}{(Fe/H)_\odot}\right) \tag{3}$$

Feature Engineering

The features used will be a combination of the mass of the planet's star and the distance of the star to the planet. The new feature will be defined as:

$$(4\pi d^2)F \propto M^{3.5} \tag{4}$$

Now, since the surface temperature can be used as a zero-order proxy for habitability, we can correlate the flux will be a suitable feature.

$$F_{habitability} = \frac{M^{3.5}}{d^2} \quad \text{where } M < 10M_E \tag{5}$$

Filtering the Data

```python

# use these columns for the final DataFrame
columns = ['P_NAME', 'S_NAME', 'S_MASS', 'S_RADIUS', 'S_TEMPERATURE', 'P_MASS', \
        'P_RADIUS', 'P_HABITABLE', 'S_METALLICITY', 'P_DISTANCE', 'P_TYPE', 'P_MASS']

# filter stars with metallicity > 0.00 (Tylium) and planets with mass < 10 Earth masses
filtered_df = bindf[bindf['S_METALLICITY'] > 0.00]
filtered_df = filtered_df[bindf['P_MASS'] < 10]

filtered_df = pd.DataFrame(filtered_df, columns=columns)

# add the new feature M^3.5 / d^2
filtered_df.loc[:, 'F_habitability'] = \
        (filtered_df['S_MASS'] ** 3.5) / (filtered_df['P_DISTANCE'] ** 2)

```

This feature will correlate with habitability because it combines the mass of the star and the distance to the planet.

Classification

Using the new feature $F_{habitability}$ and the metallicity of the star, we can classify planets as habitable or not habitable. The classification is done using a decision tree classifier. This code was provided by the included notebook, and modified for the new feature and metallicity.

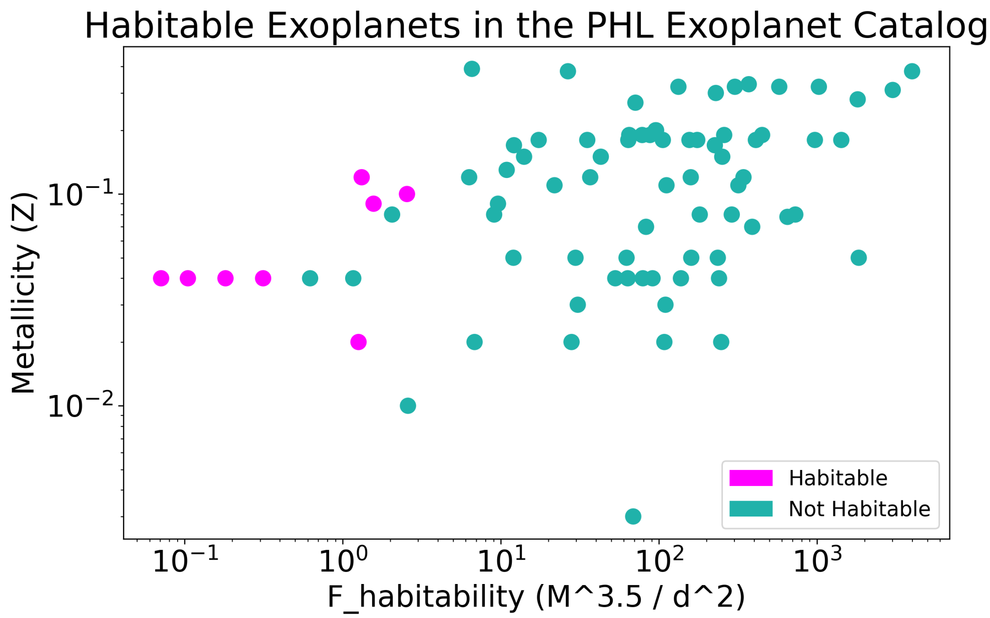

Figure 1: Feature Engineering: The new feature $F_{habitability}$ plotted against the metallicity of the star.

Decision Tree Classifier

The decision tree is constructed using the new feature $F_{habitability}$ and the metallicity of the star. The decision tree is trained to classify planets as habitable or not habitable based on these features.

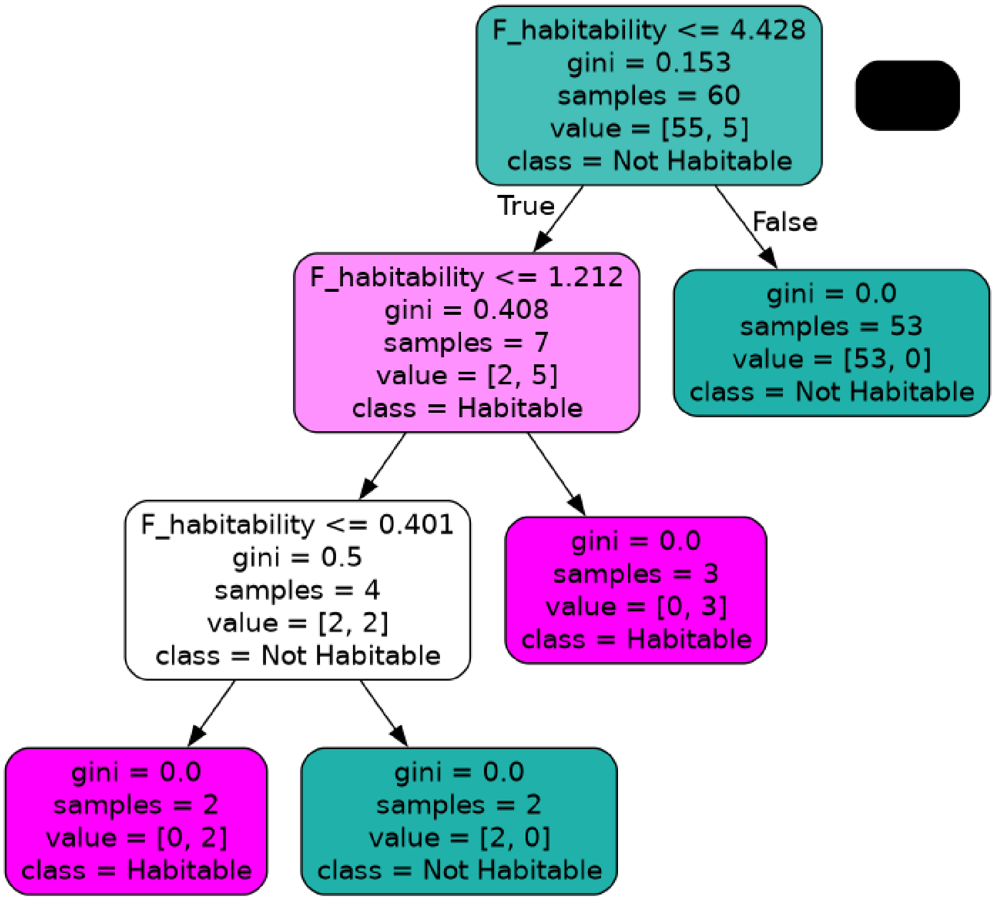

Figure 2: Decision Tree Classifier: The decision tree constructed using the new feature $F_{habitability}$ and the metallicity of the star.

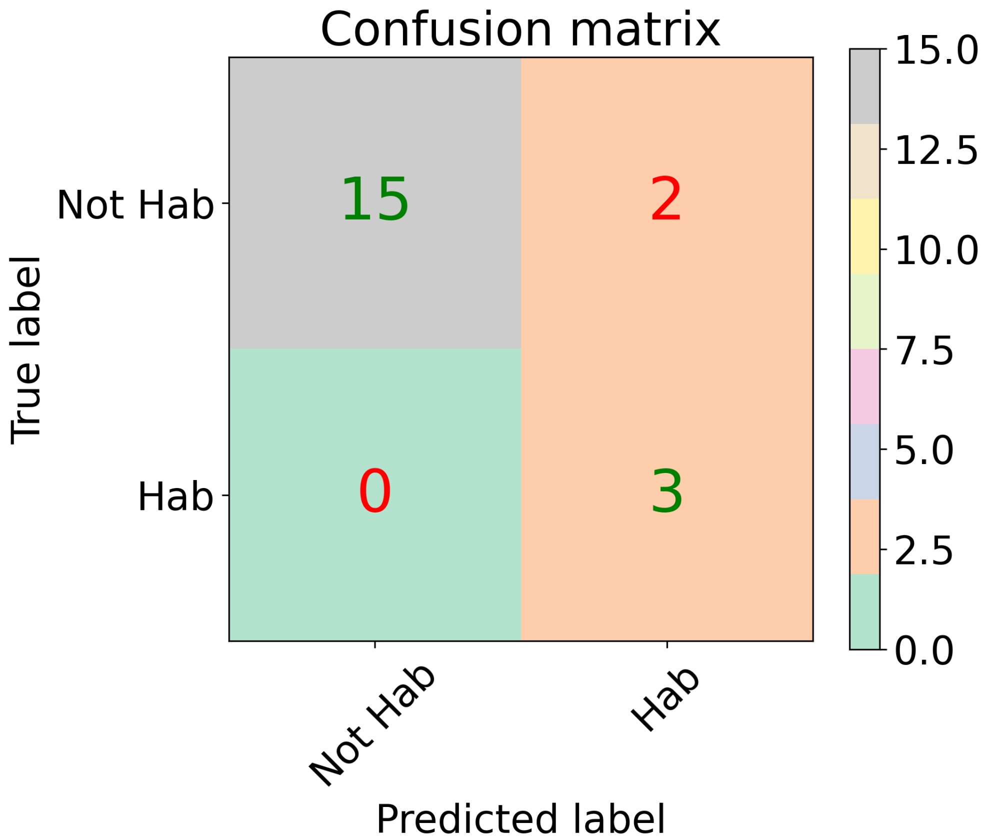

Figure 3: Confusion Matrix: The confusion matrix for the decision tree classifier.

Conclusion

Habitability can be predicted and correlates strongly with the engineered feature $F_{habitability}$, which was derived from the physical properties of the star and planet. Other constraints, even fictional ones, such as the presence of Tylium, can be used to filter the data and improve the classification. The decision tree classifier is able to classify planets as habitable or not habitable based on $F_{habitability}$ and the metallicity of the star.

References

[1] V. Acquaviva. *Machine Learning for Physics and Astronomy.* Cambridge University Press, 2025.

[2] Battlestar Galactica Wiki. Tylium, 2025. URL `https://galactica.fandom.com/wiki/Tylium`. Accessed: 2025-07-15.

[3] K. Johnson. Stellar nucleosynthesis: That's so metal, August 2020. URL `https://astrobites.org/2020/08/24/stellar-nucleosynthesis-thats-so-metal/`. Accessed: 2025-07-24.

[4] D. L. Moche. *Astronomy A Self-Teaching Guide.* Wiley, 2015.

[5] S. Weinberg. Lecture 2: Stellar structure and evolution, 2021. URL `https://www.astronomy.ohio-state.edu/weinberg.21/Intro/lec2.html`. Accessed: 2025-07-14.