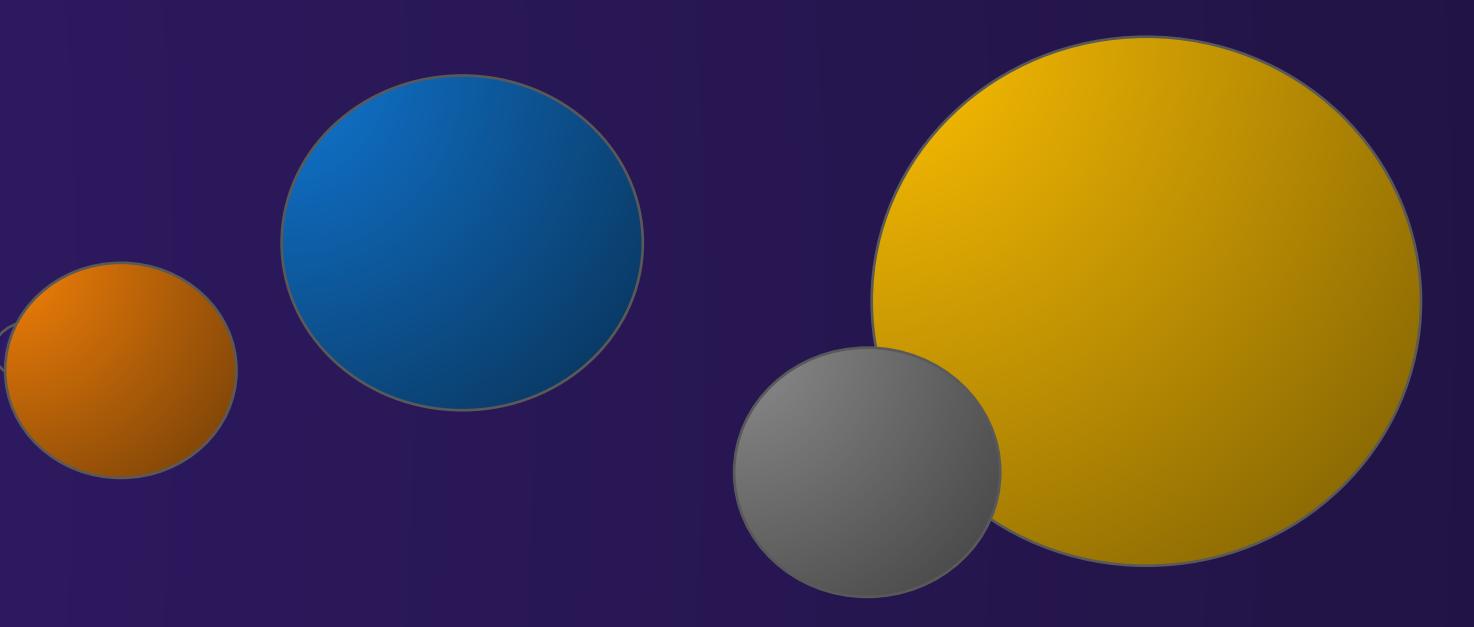




# Analyzing the Impacts of Stellar and Planetary Parameters to Exoplanet Habitability through Machine Learning

Christina X. Liu<sup>1</sup>, Jonathan H. Jiang<sup>2</sup>

<sup>1</sup>Lakeside School, Seattle, WA, <sup>2</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA



## INTRODUCTION

Are there any planets other than Earth potentially habitable by humans? What factors affect habitability and how do they impact habitability? Seeking answers to these questions has been an active research field in astrophysics in the past decade.

In this study, we identified the most influential stellar and planetary parameters to exoplanet habitability and analyzed how they impact habitability through the lens of explainable machine learning.

## OBJECTIVES

Build high-quality ML models (**Random Forest**, **XGBoost**) for exoplanet habitability prediction.

Perform analysis on the ML models through explainable AI technique **SHAP** (SHapley Additive exPlanations) to understand how stellar and planetary parameters impact exoplanet habitability.

## DATASET

The primary data sources for this study:

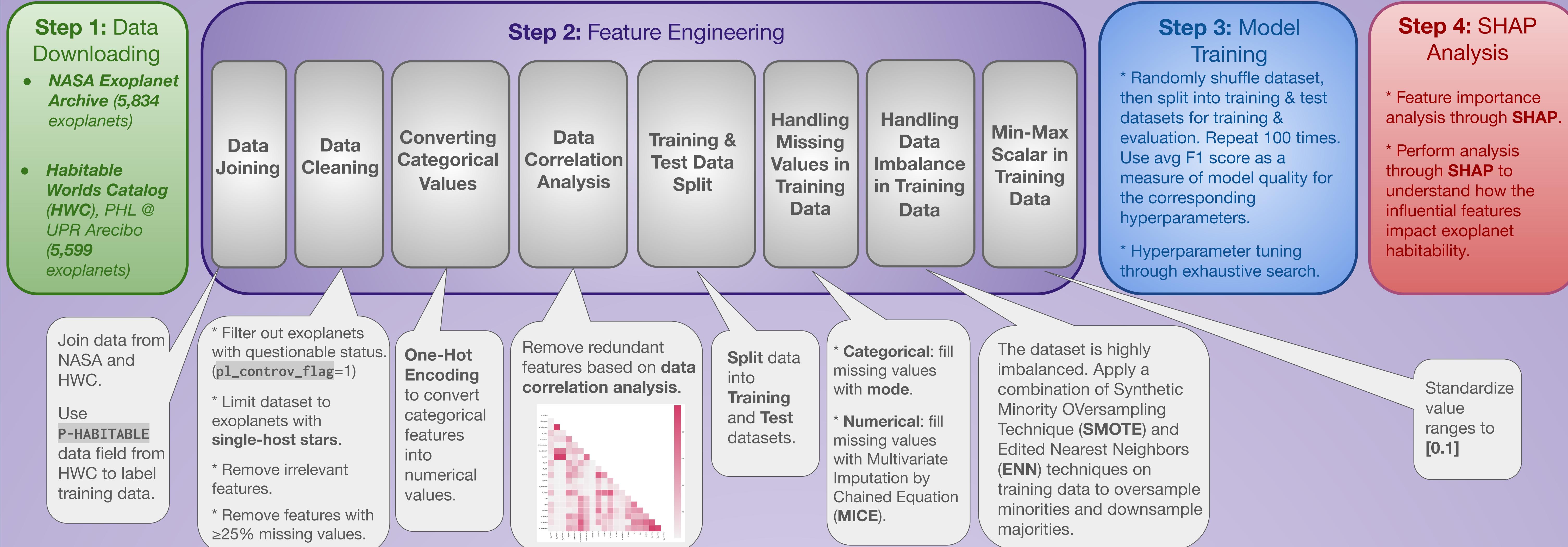
- Planetary Systems Composite Data @ **NASA Exoplanet Archive**: 5,834 confirmed exoplanets.
- Habitable Worlds Catalog (**HWC**), PHL @ UPR Arecibo: 5,599 exoplanets.

The NASA dataset was joined with the HWC dataset. Then, the **P\_HABITABLE** data field from the HWC dataset was used to label the combined data.

## MODEL SELECTION

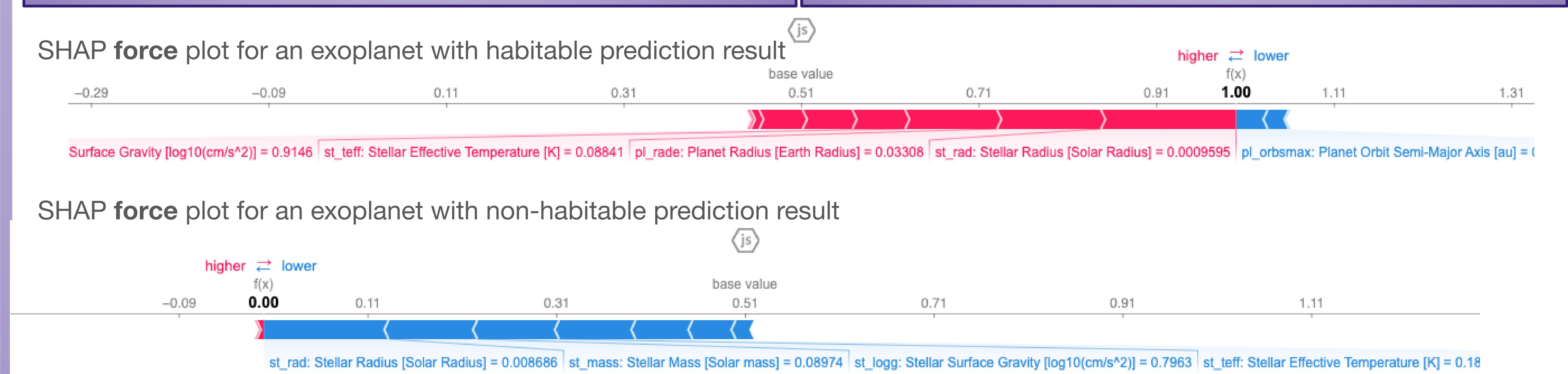
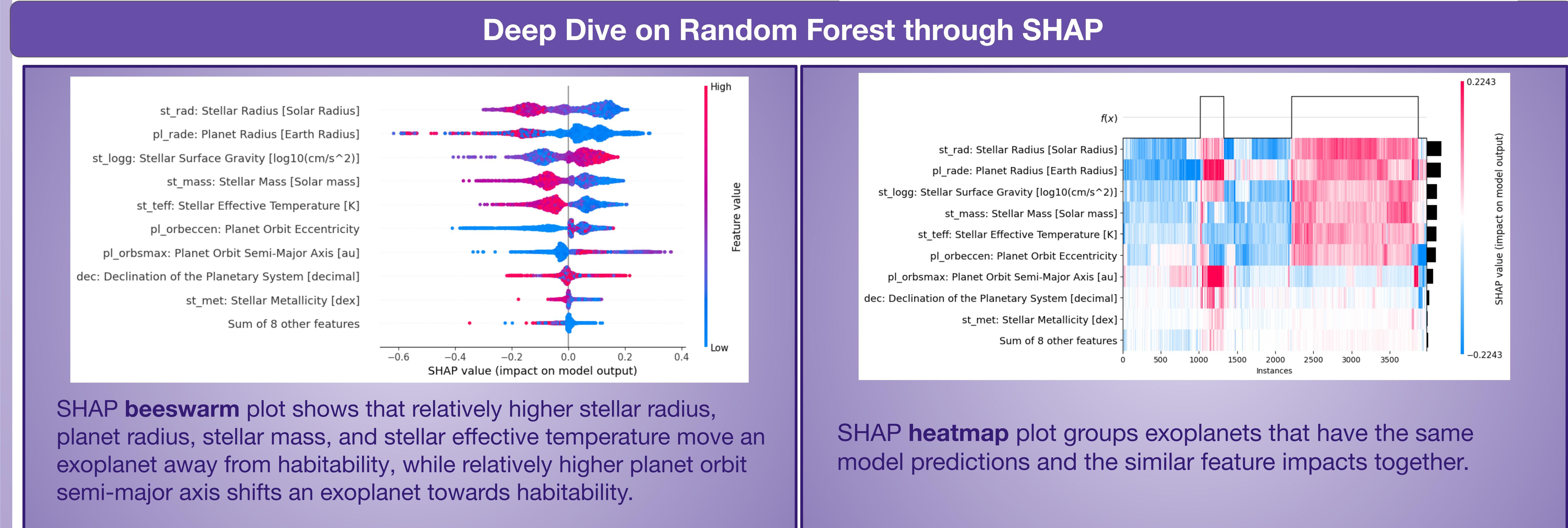
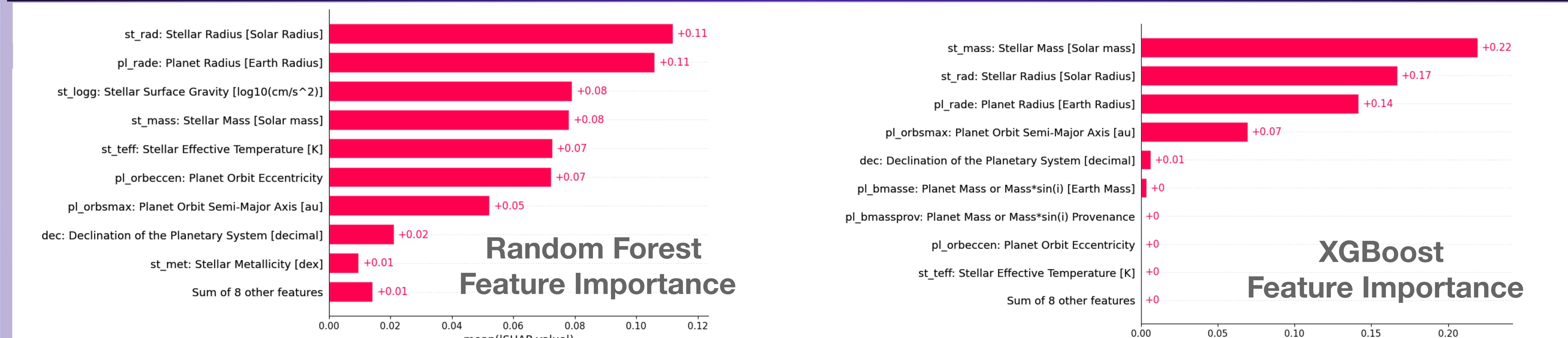
Research in machine learning shows that tree-based models can be very effective for tabular dataset (on par with deep learning models). In this study, we chose to build **Random Forest** and **XGBoost** classifiers for exoplanet habitability prediction.

## FEATURE ENGINEERING & MODEL TRAINING

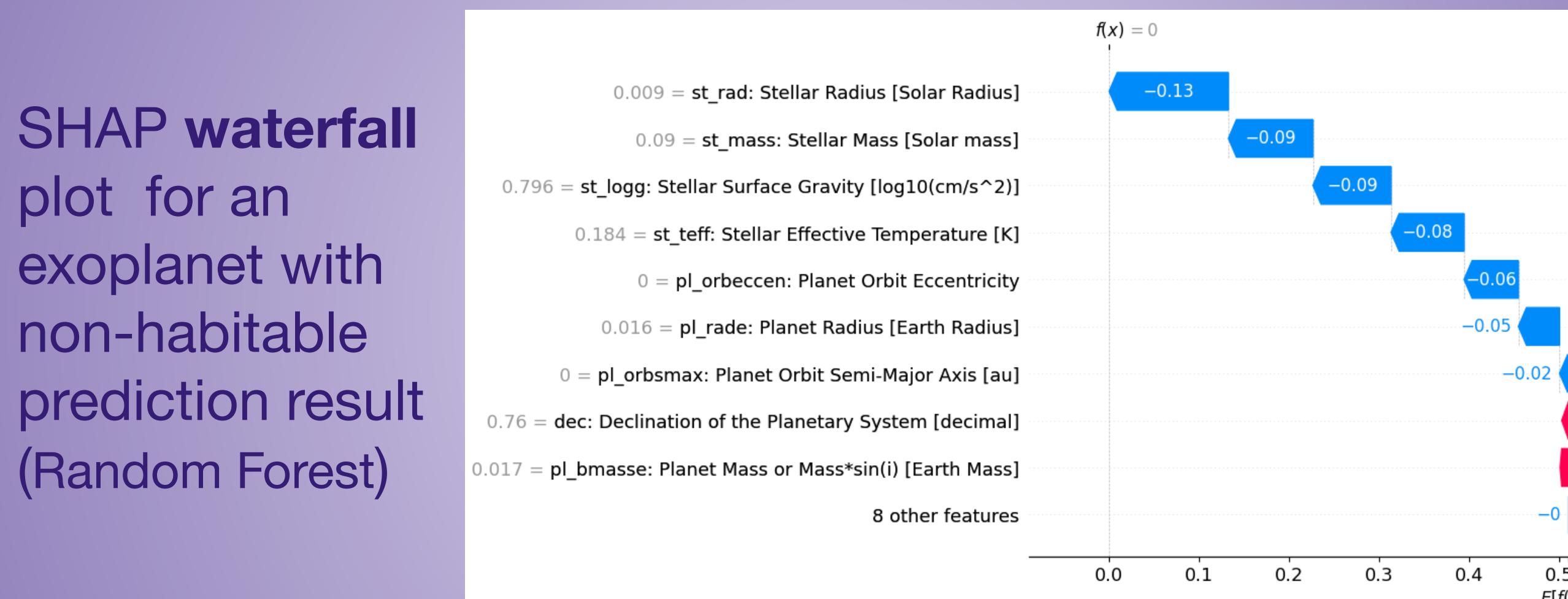
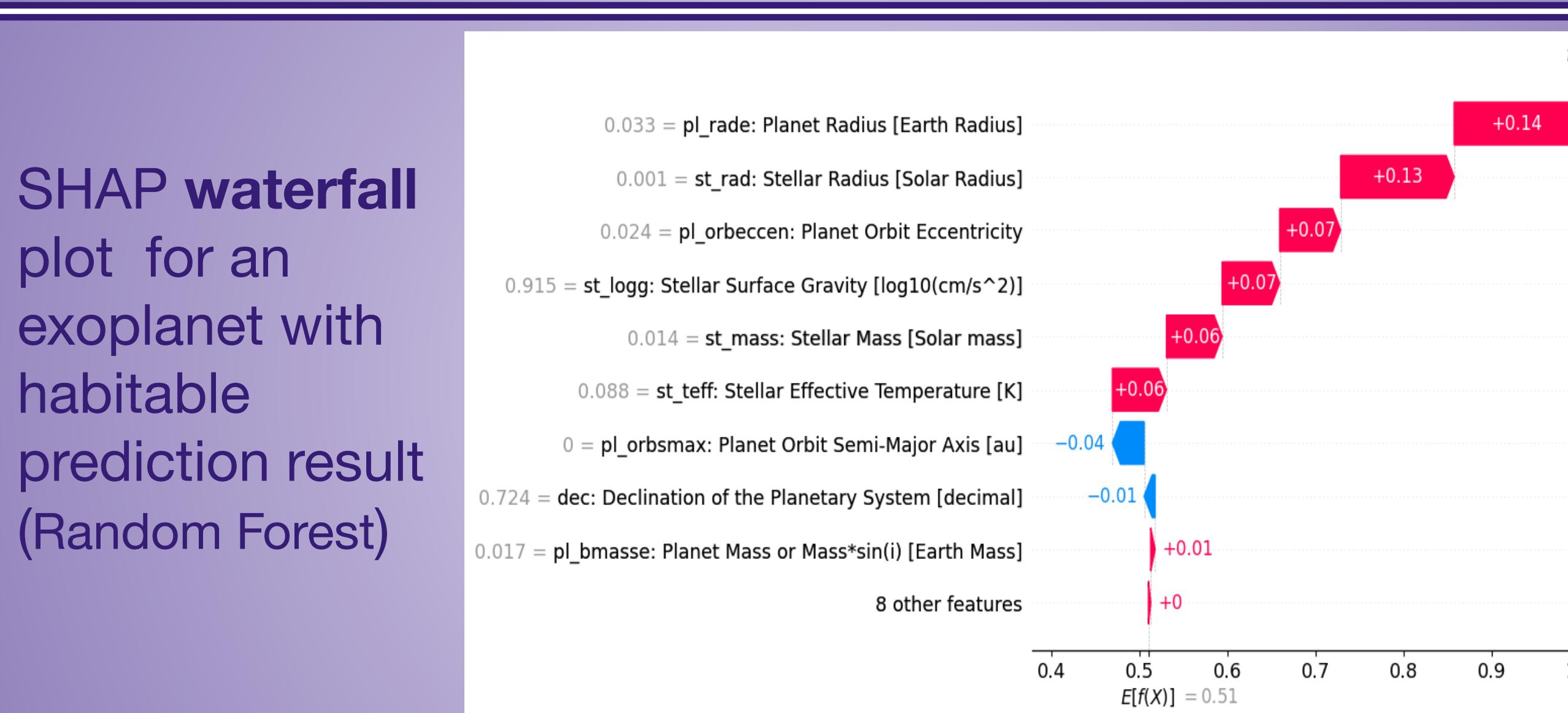
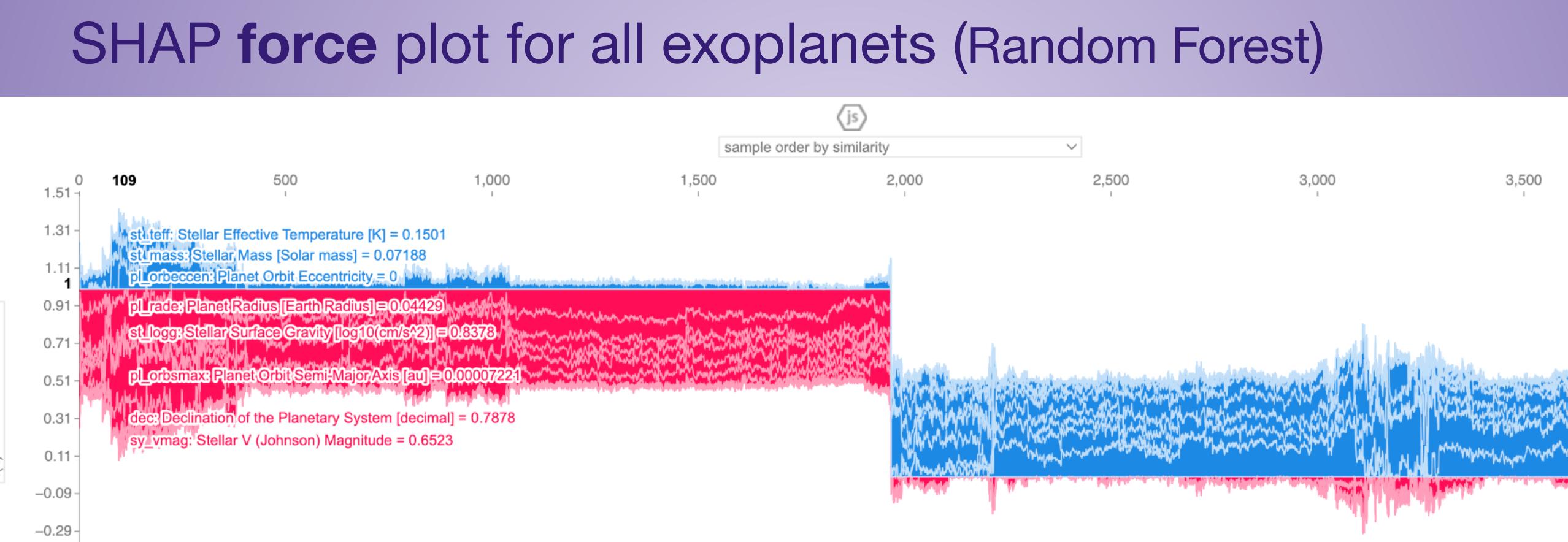


	Precision	Recall	F1 Score
<b>Random Forest Model Evaluation</b>	0.95	0.94	0.95
<b>XGBoost Model Evaluation</b>	0.94	0.97	0.95

## SHAP ANALYSIS



## MORE SHAP ANALYSIS



## CONCLUSIONS

**Random Forest** and **XGBoost** models were trained to predict exoplanet habitability with precisions at **0.95** and **0.94**.

Feature importance analysis through **SHAP** identified influential stellar and planetary parameters to habitability. Further **SHAP** analysis on these parameters showed that relatively higher stellar radius, stellar mass, stellar effective temperature, and planet radius move an exoplanet away from habitability, while relatively higher planet orbit semi-major axis shifts an exoplanet towards habitability.

This study demonstrated that explainable machine learning can be effective for both predicting exoplanet habitability and understanding how stellar and planetary parameters impact exoplanet habitability.

## FUTURE WORK

Train a **Neural Network** model for habitability prediction and **SHAP** analysis. Compare with tree-based models.

Study **Planetary Systems** as a whole to understand what planetary systems might be more likely to host habitable planets.