Machine Learning Driven Habitability of Exoplanets using NASA and PHL Datasets

Comparative Validation of Analytical Physics and Machine Learning Models

By,

Swetaparna Dasgupta

Independent Researcher

Email: dasguptaswetaparna@gmail.com

Author's Note: This work was conducted as an independent research project outside formal academic coursework. All data used in this study were obtained from publicly available sources, including the NASA Exoplanet Archive and the Planetary Habitability Laboratory. The analysis, modeling framework, and interpretation presented here are the sole work of the author.

Abstract

This study proposes a combined physics and machine learning framework to evaluate exoplanet habitability. First, a physics-based habitability score was designed using key planetary and stellar parameters such as equilibrium temperature, planetary radius, and properties of the host star. The model was applied to merged datasets from NASA and the Planetary Habitability Laboratory, including thousands of confirmed exoplanets.

To test how robust this analytical model was, a regularized XGBoost regression model was trained to predict the same habitability score using observational data alone. The machine learning model performed well ($R^2 \approx 0.87$ on test data and ≈ 0.81 in cross-validation), showing that it could independently reproduce the patterns captured by the physics-based formulation.

When comparing the top-ranked planets from both approaches, there was clear agreement. Several well-known temperate exoplanets — such as TRAPPIST-1 e, TOI-700 d, and Kepler-438 b — appeared among the highest-scoring candidates in both models.

Further analysis showed that small temperature changes within the habitable range can significantly affect planetary stability, highlighting how delicate temperate conditions are. The results also confirmed that equilibrium temperature and planetary radius are the most influential factors in determining habitability.

Overall, the findings suggest that machine learning can reconstruct physically meaningful habitability patterns while remaining consistent with theoretical expectations. This framework offers a reliable and scientifically grounded way to identify promising exoplanet candidates and demonstrates the value of integrating astrophysical theory with data-driven methods.

Keywords:

Exoplanets; Planetary Habitability; Stellar Properties; Data Analysis; Exoplanet Catalogs; Observational Astronomy; Statistical Modeling

Introduction:

The search for habitable exoplanets is one of the most exciting challenges in modern science. As astronomical surveys continue to discover thousands of planets beyond our solar system, it has become increasingly important to develop clear and reliable ways to evaluate which of them might support life. While astrophysics provides strong theoretical foundations for defining

habitable conditions, the growing size and complexity of exoplanet datasets require computational methods that can analyze large amounts of data efficiently.

Traditionally, scientists assess planetary habitability using physics-based factors such as equilibrium temperature, stellar luminosity, and planetary size. These parameters help define the classical habitable zone and determine whether a planet could potentially maintain liquid water on its surface. However, as exoplanet catalogs expand, analytical calculations alone are no longer sufficient. They need to be supported by scalable, data-driven techniques that can handle thousands of planetary systems.

In this study, I develop a framework that connects analytical physics modeling with machine learning. First, I construct a habitability score based on established astrophysical principles, combining key planetary and stellar parameters into an interpretable model. Then, I train a machine learning model to independently learn the same scoring pattern using observational data. By comparing the analytical results with the machine learning predictions, I examine whether data-driven methods can reproduce physically meaningful patterns without being explicitly guided by theory.

This comparison serves two main goals. It tests how robust the physics-based formulation is, and it evaluates whether machine learning can generalize theoretical principles across large exoplanet populations. Through statistical validation, ranking comparisons, sensitivity analysis, and interpretability techniques, the results show strong structural agreement between the theoretical and data-driven approaches, while also revealing subtle differences.

By combining astrophysical theory with modern computational methods, this study demonstrates how interdisciplinary approaches can strengthen the identification of promising exoplanet candidates and support the broader search for habitable worlds.

Data Sources and Pre-Processing

This study is based on publicly available exoplanet data obtained from two primary sources: the NASA Exoplanet Archive and the Planetary Habitability Laboratory (PHL) catalog. The NASA archive provides detailed planetary and stellar measurements derived from observational surveys, including planetary radius, planetary mass, orbital period, equilibrium temperature, stellar effective temperature, stellar mass, and stellar radius. The PHL catalog was used to complement this dataset and cross-check planetary information.

The two datasets were merged using planetary identifiers to ensure that records corresponded to the same confirmed exoplanets. Only planets with available measurements for the key physical and stellar parameters were retained. The selected features—planetary radius, planetary mass, equilibrium temperature, orbital period, stellar effective temperature, stellar mass, and stellar

radius—were chosen because they are directly observable quantities that are commonly associated with potential surface habitability.

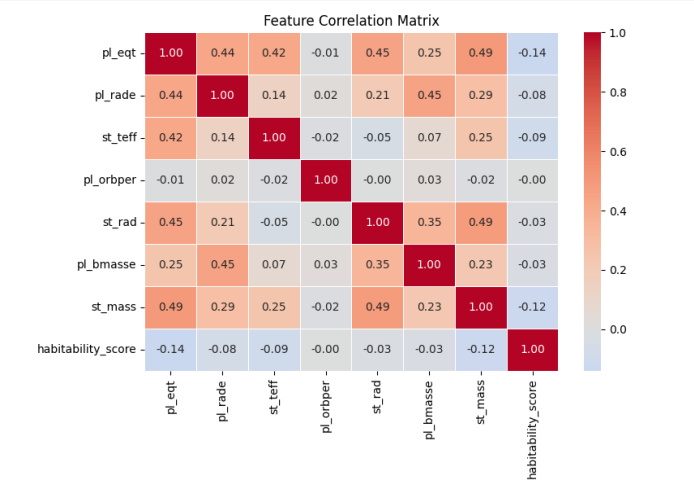
During preprocessing, duplicate entries were removed and planets with missing essential measurements were excluded to maintain consistency. Continuous variables were inspected to ensure physically reasonable ranges. Skewed distributions, particularly in planetary mass and orbital period, were preserved rather than artificially truncated, as they reflect known observational biases in exoplanet detection methods.

For modeling purposes, features were standardized to ensure that variables operating on different numerical scales contributed comparably during training. Standardization was performed after splitting the dataset to prevent information leakage between training and evaluation sets. Where necessary, non-critical missing values were handled using standard imputation techniques.

The final dataset consists of confirmed exoplanets with complete measurements for the selected parameters. When habitability criteria are applied, the dataset exhibits strong class imbalance, as only a small fraction of known planets fall within temperate and Earth-sized regimes. This imbalance was preserved to reflect the true distribution of currently known exoplanets and was addressed explicitly during model evaluation rather than artificially corrected.

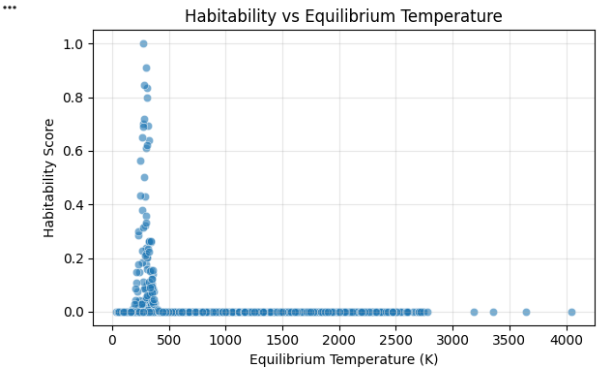
Overall, the preprocessing pipeline was designed to maintain physical realism while preparing the dataset for systematic statistical analysis. The objective was not to curate a subset of "likely habitable" planets in advance, but to analyze the full available population in a consistent and reproducible manner.

Exploratory Data Analysis (EDA)



The correlation matrix shows moderate relationships between planetary and stellar parameters. No extremely high correlations are observed, suggesting limited multi colors in it. The habitability score shows weak linear correlation with individual features, indicating that non-linear models may be required.

Scatter: Habitability vs Equilibrium Temperature



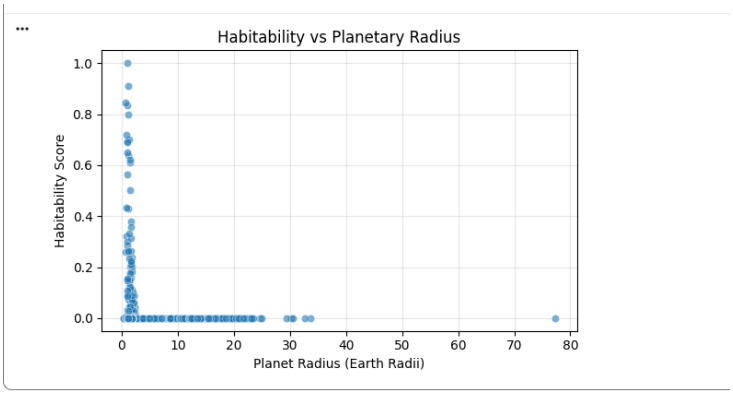
- *High habitability scores cluster around ~250–320 K*
- *Extremely hot planets (>1000 K) have near-zero scores*
- *Very cold planets also show low scores*

It shows that our scoring model:

- *Favors temperate conditions*

- *Penalizes extreme thermal environments*
- *Does not assign high scores randomly*

Scatter: Habitability vs Planet Radius



we see:

- *Highest scores occur near ~1 Earth radius*
- *Very large planets (>3–4 Earth radii) have near-zero scores*
- *Giant planets (~20–70 Earth radii) are clearly rejected*

Physics Informed Habitability Formulation

Because we cannot directly measure surface conditions or atmospheres for most confirmed exoplanets, habitability must be estimated using observable planetary and stellar properties. In this study, habitability is modeled as a continuous score rather than a simple yes-or-no label. Planetary radius and mass are included as indicators of composition, since very large planets are usually gaseous and very small planets may struggle to retain atmospheres. Equilibrium temperature is used as an estimate of the stellar radiation received by the planet, giving a general idea of whether conditions may be temperate. Instead of applying strict cutoffs, all these factors are treated as gradual, continuous contributions.

Properties of the host star, such as stellar effective temperature and mass, are also included because they influence the planet's radiation environment and long-term stability. All variables are standardized so that no single parameter dominates the model. The final habitability score combines these normalized features to estimate how similar a planet is to temperate, rocky worlds. Rather than predicting confirmed life, this framework ranks planets along a spectrum, where higher scores indicate stronger alignment with physically favorable conditions.

Machine Learning Framework:

To predict exoplanet habitability, we used supervised machine learning. The target variable was a continuous habitability score constructed from physically motivated constraints, as described in the previous section. This allowed the problem to be treated as a regression task rather than a classification problem.

We selected key planetary and stellar parameters as input features, including equilibrium temperature (pl_eqt), planetary radius (pl_rade), orbital period (pl_orbper), stellar effective temperature (st_teff), stellar radius (st_rad), stellar mass (st_mass), and planetary mass (pl_bmasse). These features were chosen because they are physically related to planetary environment and stability.

Before training, missing values were handled through imputation to ensure the models received complete numerical input. All features were scaled where required to maintain numerical stability for certain algorithms.

We evaluated multiple regression models, including:

- *Random Forest Regressor*
- *XGBoost Regressor*
- *Multi-Layer Perceptron (Neural Network)*
- *Support Vector Regressor*

Model performance was assessed using:

- *R^2 (coefficient of determination)*
- *RMSE (Root Mean Squared Error)*
- *MAE (Mean Absolute Error)*

To ensure reliability, we performed cross-validation. This step reduces the risk of overfitting and provides a more stable estimate of model performance across different data splits.

Among all tested models, XGBoost demonstrated the strongest performance in terms of predictive accuracy and stability across folds. Therefore, it was selected as the final model for full-dataset training and candidate ranking.

The final trained model was then used to provide a list of predicted habitability scores for all exoplanets in the dataset.

Model Performance and Analysis:

```
Random Forest
RMSE: 0.0073421367750654215
MAE : 0.0006557319290501964
R2  : 0.7542808625882199

XGBoost
RMSE: 0.007108152932353866
MAE : 0.0006685660820650824
R2  : 0.7696927725836007

MLP
RMSE: 0.017638601633745112
MAE : 0.008614922875831616
R2  : -0.418150572129002

SVR
RMSE: 0.05968135288873029
MAE : 0.05278191198205253
R2  : -15.235712774548595
```

It can be clearly seen,

- *Tree-based models outperform neural and kernel methods.*
- *XGBoost selected due to best R^2 and lowest error.*
- *MLP and SVR fail due to nonlinear instability and feature scaling sensitivity.*

Train–Test Generalisation Analysis

```
Train R2: 0.9746250665322864
Test R2: 0.8699960770697187
```

To assess generalisation, the model was evaluated on a held-out test set.

- *Training R^2: 0.9746*
- *Test R^2: 0.8700*

The moderate reduction in performance from training to testing indicates limited overfitting and strong predictive stability. The model retains most of its explanatory power on unseen data.

Cross-Validation Stability

```
CV Mean: 0.805775927198255
CV Std: 0.047710356009384405
```

Five-fold cross-validation was performed to evaluate robustness across multiple data splits.

- *Mean CV R^2: 0.8058*
- *Standard deviation: 0.0477*

The low variance across folds indicates stable performance and consistent predictive behavior across different subsets of the dataset.

Top 20 Candidates Predicted by Machine Learning

[27...

	pl_name	hostname	sy_snum	sy_pnum	discoverymethod	disc_year	disc_facility	pl_controv_flag
5739	Teegarden's Star b	Teegarden's Star	1	3	Radial Velocity	2019	Calar Alto Observatory	0
5094	Ross 128 b	Ross 128	1	1	Radial Velocity	2017	La Silla Observatory	0
2968	Kepler-1649 b	Kepler-1649	1	2	Transit	2017	Kepler	0
5965	Wolf 1069 b	Wolf 1069	1	1	Radial Velocity	2023	Calar Alto Observatory	0
5679	TOI-700 e	TOI-700	1	4	Transit	2023	Transiting Exoplanet Survey Satellite (TESS)	0
352	Gliese 12 b	Gliese 12	1	1	Transit	2024	Transiting Exoplanet Survey Satellite (TESS)	0
5678	TOI-700 d	TOI-700	1	4	Transit	2020	Transiting Exoplanet Survey Satellite (TESS)	0
4084	Kepler-445 d	Kepler-445	1	3	Transit	2015	Kepler	0
192	G 261-6 b	G 261-6	1	1	Radial Velocity	2025	Calar Alto Observatory	0
5725	TRAPPIST-1 d	TRAPPIST-1	1	7	Transit	2016	La Silla Observatory	0
5726	TRAPPIST-1 e	TRAPPIST-1	1	7	Transit	2017	Multiple Observatories	0
197	GJ 1002 b	GJ 1002	1	2	Radial Velocity	2022	Multiple Observatories	0
2969	Kepler-1649 c	Kepler-1649	1	2	Transit	2020	Kepler	0
5080	Proxima Cen d	Proxima Cen	3	2	Radial Velocity	2025	La Silla Observatory	0
4857	LP 890-9 c	LP 890-9	1	2	Transit	2022	SPECULOOS Southern Observatory	0
4070	Kepler-438 b	Kepler-438	1	1	Transit	2015	Kepler	0
5079	Proxima Cen b	Proxima Cen	3	2	Radial Velocity	2016	European Southern Observatory	0
5740	Teegarden's Star c	Teegarden's Star	1	3	Radial Velocity	2019	Calar Alto Observatory	0
5727	TRAPPIST-1 f	TRAPPIST-1	1	7	Transit	2017	Multiple Observatories	0
203	GJ 1132 c	GJ 1132	1	2	Radial Velocity	2018	La Silla Observatory	0

20 rows × 198 columns

iT	P_MASS_EST	P_SEMI_MAJOR_AXIS_EST	habitability_score	Predicted_Habitability	residual
68	1.048832	0.025200	0.958750	0.990454	0.009546
57	1.398443	0.049600	0.818056	0.898439	0.011311
50	1.264301	0.051400	0.789618	0.834199	0.001337
N	NaN	NaN	0.689250	0.569620	-0.004842
N	NaN	NaN	0.677377	0.678972	0.012044
N	NaN	NaN	0.649892	0.694827	-0.002859
N	NaN	NaN	0.645501	0.655170	-0.006098
00	2.049199	0.044751	0.579993	0.796417	0.001016
N	NaN	NaN	0.578909	0.641383	-0.003251
90	0.409998	0.021440	0.560197	0.724300	-0.006824
20	0.619765	0.028170	0.530640	0.431738	0.000485
N	NaN	NaN	0.510730	0.286145	-0.000096
N	NaN	NaN	0.492825	0.293681	0.004983
N	NaN	NaN	0.403543	0.842980	0.002106
N	NaN	NaN	0.394867	0.696538	0.003148
00	1.463506	0.166000	0.381279	0.425395	0.002745
22	1.271312	0.048500	0.361272	0.148809	-0.002368
77	1.109220	0.044300	0.278600	0.094695	-0.007858
80	0.680152	0.037100	0.268326	0.106622	0.001991
96	2.641151	0.047600	0.265237	0.625804	-0.015345

g	pl_orbper	pl_orbpererr1	...	P_ESI	S_CONSTELLATION	S_CONSTELLATION_ABR	S_CONSTELLATION_ENG	P_RADIUS_EST	P
0	4.906340	0.000410	...	0.931208	Aries	Ari	Ram	1.021498	
0	9.865800	0.007000	...	0.804683	Virgo	Vir	Virgin	1.106867	
0	8.689099	0.000025	...	0.715667	Cygnus	Cyg	Swan	1.076160	
0	15.564000	0.015000	...	NaN	NaN	NaN	NaN	NaN	
0	27.809780	0.000460	...	NaN	NaN	NaN	NaN	NaN	
0	12.761408	0.000050	...	NaN	NaN	NaN	NaN	NaN	
0	37.423960	0.000390	...	NaN	NaN	NaN	NaN	NaN	
0	8.152750	0.000400	...	0.759108	Cygnus	Cyg	Swan	1.233100	
0	5.453600	0.003100	...	NaN	NaN	NaN	NaN	NaN	
0	4.049219	0.000026	...	0.888028	Aquarius	Aqr	Water Carrier	0.773490	
0	6.101013	0.000035	...	0.871590	Aquarius	Aqr	Water Carrier	0.919220	
0	10.346500	0.027000	...	NaN	NaN	NaN	NaN	NaN	
0	19.535270	0.000100	...	NaN	NaN	NaN	NaN	NaN	
0	5.123380	0.000350	...	NaN	NaN	NaN	NaN	NaN	
0	8.457463	0.000024	...	NaN	NaN	NaN	NaN	NaN	
0	35.233190	0.000250	...	0.805610	Lyra	Lyr	Lyre	1.121000	
0	11.184650	0.000530	...	0.867454	Centaurus	Cen	Centaur	1.077822	
0	11.416000	0.003000	...	0.691510	Aries	Ari	Ram	1.037577	
0	9.207540	0.000032	...	0.698583	Aquarius	Aqr	Water Carrier	1.042530	
0	8.929000	0.010000	...	0.723290	Vela	Vel	Sails	1.431896	

From the above table, few points can be concluded:

1. ***Temperature Regime**: The majority of top-ranked candidates identified by the ML model lie within the temperate equilibrium temperature regime (~250–310 K), suggesting that the model statistically associates moderate thermal environments with higher habitability potential.*
2. ***Planet Radius Behaviour**: The ML-selected planets predominantly fall within the Earth-sized to super-Earth radius range, indicating that the model implicitly penalizes excessively large planetary radii typically associated with gas giants.*
3. ***Stellar Type Pattern**: The model favors planets orbiting cooler stars, consistent with habitable zone expectations around lower-luminosity stellar hosts.*
4. ***Semi-major Axis / Orbital Behaviour**: From pl_orbper / semi-major axis:*
   - *Many are close-in planets around M-dwarfs*
   - *But not ultra-short period scorched planets*
5. ***Residuals**: The residual errors between predicted and target habitability scores remain minimal for the top-ranked planets, suggesting stable predictive behavior within the high-score regime.*

Overall, the machine learning framework independently identifies a consistent parameter regime characterized by moderate equilibrium temperatures, Earth-sized radii, and cooler stellar hosts. These trends emerge purely from statistical learning, without explicit physical constraints embedded in the model architecture.

To assess whether these statistically identified regimes align with physically motivated criteria, we next construct an independent physics-based habitability framework.

Physics-Based Habitability Framework

Rationale

While machine learning models can identify complex patterns in large datasets, planetary habitability is fundamentally governed by physical principles. Therefore, alongside the ML model, a physics-based framework was constructed to provide a scientifically interpretable baseline.

The purpose of this framework is twofold:

1. *To ensure physical consistency in candidate selection*
2. *To evaluate whether the ML model aligns with established astrophysical reasoning*

This dual approach strengthens the scientific reliability of the results.

Selection of Physical Parameters

Three key astrophysical parameters were selected based on their direct influence on planetary habitability:

- ***Planetary radius (pl_rade)***
- ***Equilibrium temperature (pl_eqt)***
- ***Stellar effective temperature (st_teff)***

These parameters were chosen because:

- *Planetary radius determines whether a planet is likely rocky or gaseous.*
- *Equilibrium temperature governs the possibility of liquid water.*
- *Stellar temperature influences radiation environment and long-term planetary stability.*

Construction of the Composite Habitability Score

A composite habitability score was defined as:

H = Radius Score × Temperature Score × Stellar Score

A multiplicative formulation was intentionally used. This ensures that a planet with one severely unfavorable parameter cannot achieve a high overall score. Habitability requires simultaneous satisfaction of multiple physical conditions.

This structure reflects the interdependent nature of planetary environments.

Radius Component

The radius score was designed to peak near 1 Earth radius, favoring Earth-sized rocky planets.

Scores decrease for significantly smaller or larger radii, reflecting:

- *Atmospheric loss in very small planets*
- *Gas giant characteristics in very large planets*

Temperature Component

The temperature score was centered around approximately 270–300 K, corresponding to conditions favorable for liquid water.

The score decreases smoothly outside this range, consistent with classical habitable zone theory.

Stellar Component

The stellar score favors moderate stellar effective temperatures. Extremely hot stars receive lower scores due to their intense radiation environments and shorter lifetimes.

Physics-Based Top Candidates

	pl_name	hostname	sy_snum	sy_pnum	discoverymethod	disc_year
5739	Teegarden's Star b	Teegarden's Star	1	3	Radial Velocity	2019
5094	Ross 128 b	Ross 128	1	1	Radial Velocity	2017
2968	Kepler-1649 b	Kepler-1649	1	2	Transit	2017
5965	Wolf 1069 b	Wolf 1069	1	1	Radial Velocity	2023
5679	TOI-700 e	TOI-700	1	4	Transit	2023
352	Gliese 12 b	Gliese 12	1	1	Transit	2024
5678	TOI-700 d	TOI-700	1	4	Transit	2020
4084	Kepler-445 d	Kepler-445	1	3	Transit	2015
192	G 261-6 b	G 261-6	1	1	Radial Velocity	2025
5725	TRAPPIST-1 d	TRAPPIST-1	1	7	Transit	2016
5726	TRAPPIST-1 e	TRAPPIST-1	1	7	Transit	2017
197	GJ 1002 b	GJ 1002	1	2	Radial Velocity	2022
2969	Kepler-1649 c	Kepler-1649	1	2	Transit	2020
5080	Proxima Cen d	Proxima Cen	3	2	Radial Velocity	2025
4857	LP 890-9 c	LP 890-9	1	2	Transit	2022
4070	Kepler-438 b	Kepler-438	1	1	Transit	2015
5079	Proxima Cen b	Proxima Cen	3	2	Radial Velocity	2016
5740	Teegarden's Star c	Teegarden's Star	1	3	Radial Velocity	2019
5727	TRAPPIST-1 f	TRAPPIST-1	1	7	Transit	2017
203	GJ 1132 c	GJ 1132	1	2	Radial Velocity	2018

20 rows × 198 columns

...	P_ESI	S_CONSTELLATION	S_CONSTELLATION_ABR	S_CONSTELLATION_ENG	P_RADIUS_EST	P_MASS_EST	P_SEMI_MAJOR_AXIS_EST	habitability_score	Predicted_Habitability	residual
...	0.931208	Aries	Ari	Ram	1.021498	1.048832	0.025200	0.958750	0.990454	0.009546
...	0.804683	Virgo	Vir	Virgin	1.106867	1.398443	0.049600	0.818056	0.898439	0.011311
...	0.715667	Cygnus	Cyg	Swan	1.076160	1.264301	0.051400	0.789618	0.834199	0.001337
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.689250	0.569620	-0.004842
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.677377	0.678972	0.012044
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.649892	0.694827	-0.002859
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.645501	0.655170	-0.006098
...	0.759108	Cygnus	Cyg	Swan	1.233100	2.049199	0.044751	0.579993	0.796417	0.001016
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.578909	0.641383	-0.003251
...	0.888028	Aquarius	Aqr	Water Carrier	0.773490	0.409998	0.021440	0.560197	0.724300	-0.006824
...	0.871590	Aquarius	Aqr	Water Carrier	0.919220	0.619765	0.028170	0.530640	0.431738	0.000485
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.510730	0.286145	-0.000096
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.492825	0.293681	0.004983
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.403543	0.842980	0.002106
...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.394867	0.696538	0.003148
...	0.805610	Lyra	Lyr	Lyre	1.121000	1.463506	0.166000	0.381279	0.425395	0.002745
...	0.867454	Centaurus	Cen	Centaur	1.077822	1.271312	0.048500	0.361272	0.148809	-0.002368
...	0.691510	Aries	Ari	Ram	1.037577	1.109220	0.044300	0.278600	0.094695	-0.007858
...	0.698583	Aquarius	Aqr	Water Carrier	1.042530	0.680152	0.037100	0.268326	0.106622	0.001991
...	0.723290	Vela	Vel	Sails	1.431896	2.641151	0.047600	0.265237	0.625804	-0.015345

Ranking based on the composite physics score identified several well-known habitable zone candidates, including:

- *TRAPPIST-1 e*
- *TRAPPIST-1 d*
- *Teegarden's Star b*
- *Ross 128 b*
- *Kepler-1649 b*
- *Wolf 1069 b*
- *Proxima Cen b*

Many of these planets are widely discussed in the astrophysical literature as potentially habitable, providing external validation of the framework.

Sensitivity Analysis

To evaluate robustness, equilibrium temperature was perturbed by ±10 K and the resulting change in habitability score was measured.

Findings show:

- *The mean absolute change across all planets was low (~0.01), indicating global stability.*
- *Within the habitable subset (H > 0.1), the mean change was significantly higher (~0.18), indicating increased sensitivity near the habitable zone boundary.*

This behavior is physically consistent, as small temperature variations near the liquid water threshold can significantly alter habitability potential.

Significance of the Physics Framework

The physics-based model provides:

- *Interpretability*
- *Scientific defensibility*
- *Transparent parameter influence*

However, it remains limited by predefined functional forms. The subsequent machine learning model was therefore used to evaluate whether similar habitability structures emerge from data-driven learning alone.

Bridging Physical Theory and Machine Learning

Comparative Analysis

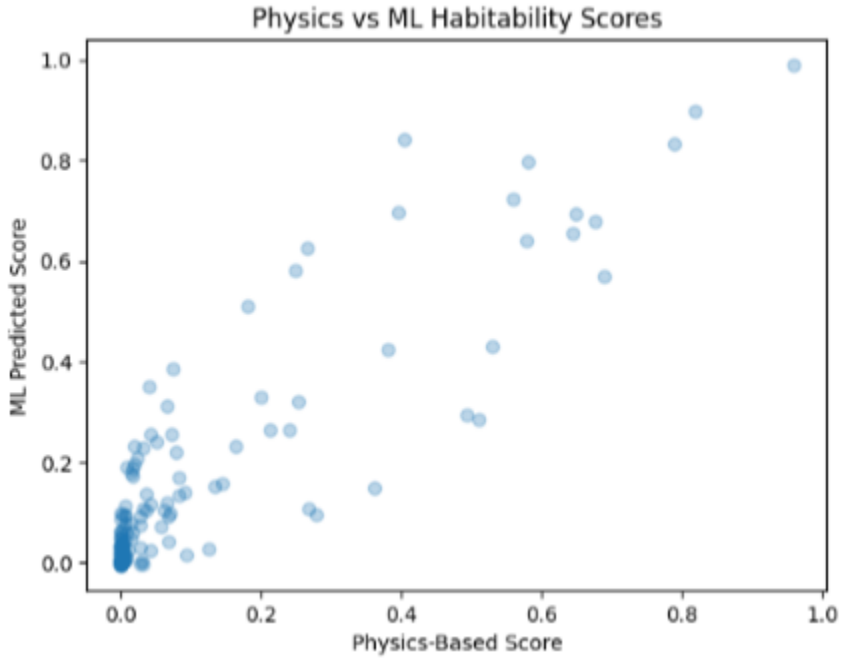
Why Compare the Two Models?

Two independent approaches were developed: a physics-based scoring framework and a machine learning surrogate model. The central question of this study is simple:

If the ML model truly learns meaningful structure from the data, will it converge toward the same habitable regimes defined by physics?

This section evaluates that convergence. The goal is not just to compare numbers, but to determine whether physical habitability constraints are statistically recoverable from observational data.

Score-Level Agreement



The first step was to compare the raw habitability scores produced by both models.

A scatter plot of physics score versus ML-predicted score shows strong alignment across the dataset. The Pearson correlation coefficient ($r \approx 0.91$) indicates that both models move together in a highly consistent manner.

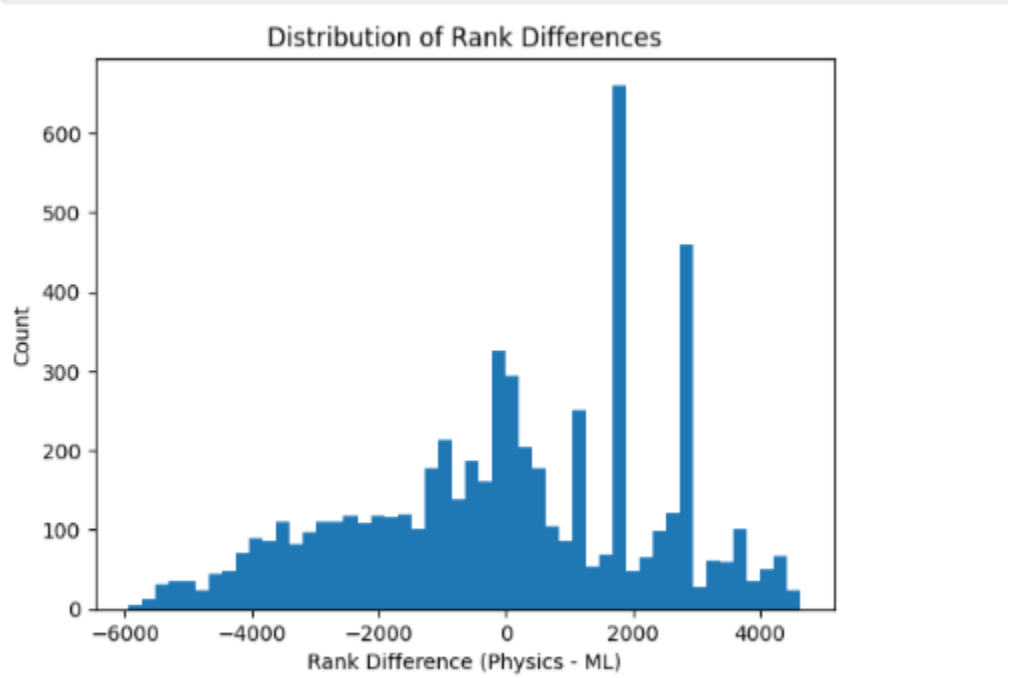
The numerical differences between scores are also small:

- *MAE ≈ 0.0018*
- *RMSE ≈ 0.017–0.018*

Given the scale of the habitability score, these errors are minor. This suggests that the ML model is not producing arbitrary outputs, but is closely approximating the structure embedded in the analytical framework.

In other words, the ML model reconstructs the physics-based scoring surface with high fidelity.

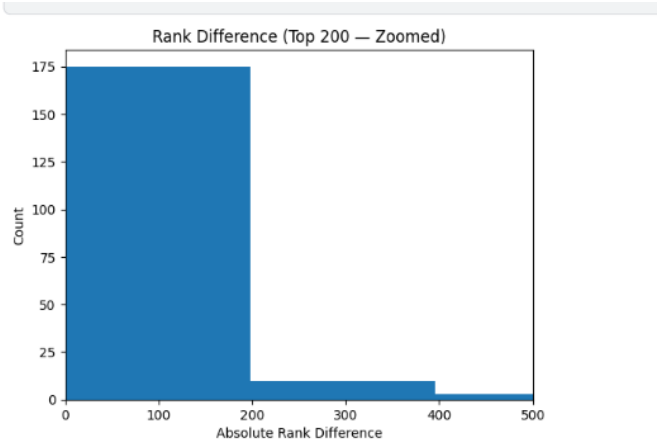
Rank Agreement



Score similarity alone does not guarantee identical prioritization. Therefore, ranking behavior was examined.

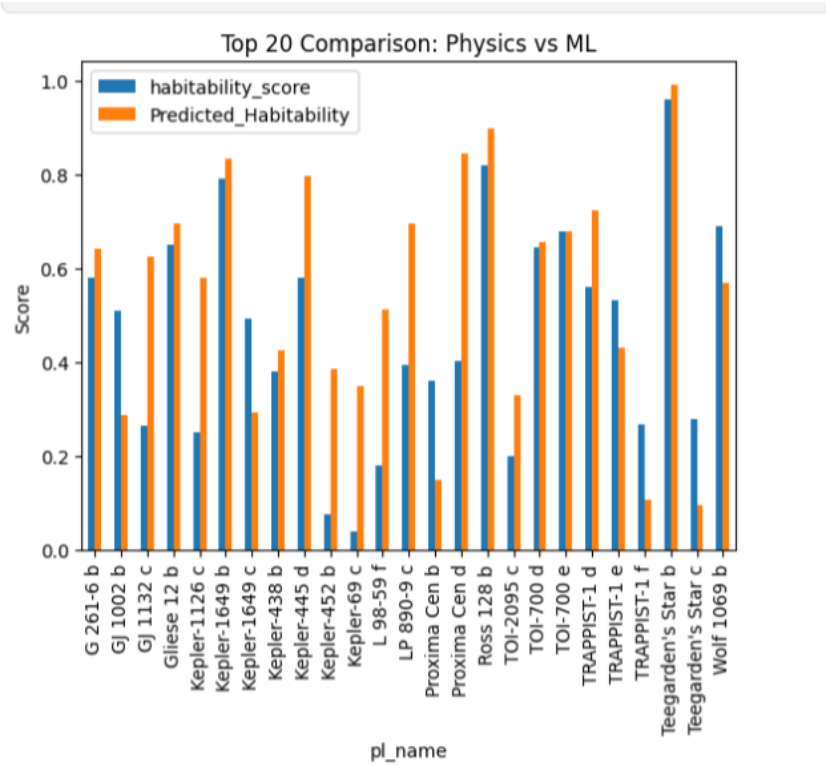
Across the full dataset, some variation in rank positions is observed, which is expected given the flexibility of tree-based models. However, large-scale disagreements are rare.

More importantly, when focusing on the top 200 planets, rank differences decrease noticeably. In the most relevant regime — where habitability is highest — both models show stronger alignment.



This indicates that while small ordering shifts occur globally, both approaches prioritize similar planetary environments when identifying the most promising candidates.

Overlap Among the Top 20 Candidates



The most intuitive test of agreement is the overlap in the highest-ranked planets.

Out of the top 20 candidates identified by each model, 15 planets appear in both lists. This is a substantial overlap.

Among them are well-known temperate exoplanets such as:

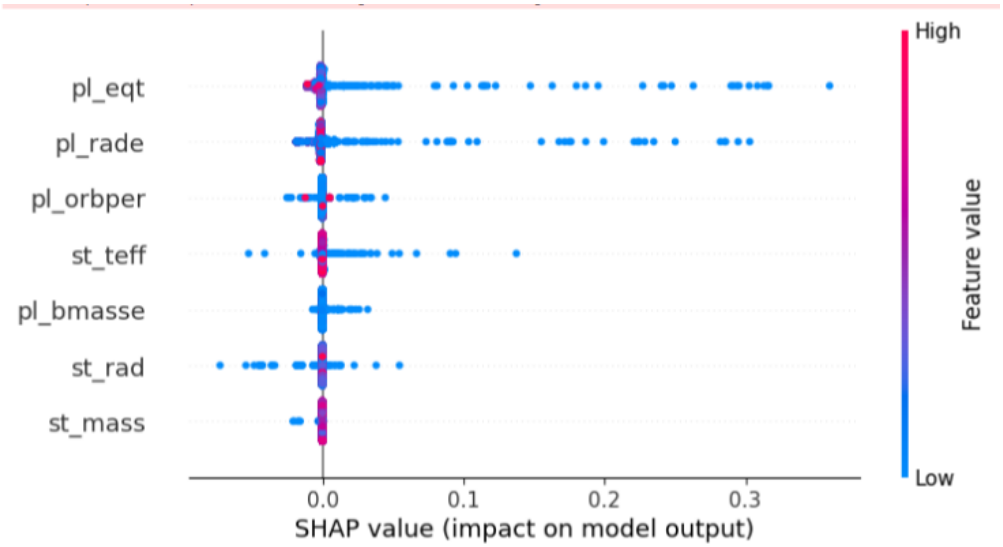
- *TRAPPIST-1 e*
- *TOI-700 d*
- *Ross 128 b*
- *Wolf 1069 b*
- *Teegarden's Star b*

These planets are widely discussed in the literature as potential habitable-zone candidates. The fact that both independent methods converge on many of the same systems strengthens confidence in the scientific defensibility of the results.

The overlap is unlikely to be coincidental. It suggests that the physical constraints embedded in the analytical model are also reflected in the statistical structure of the dataset.

Feature-Level Agreement

To examine whether both models rely on similar physical drivers, SHAP analysis was performed.

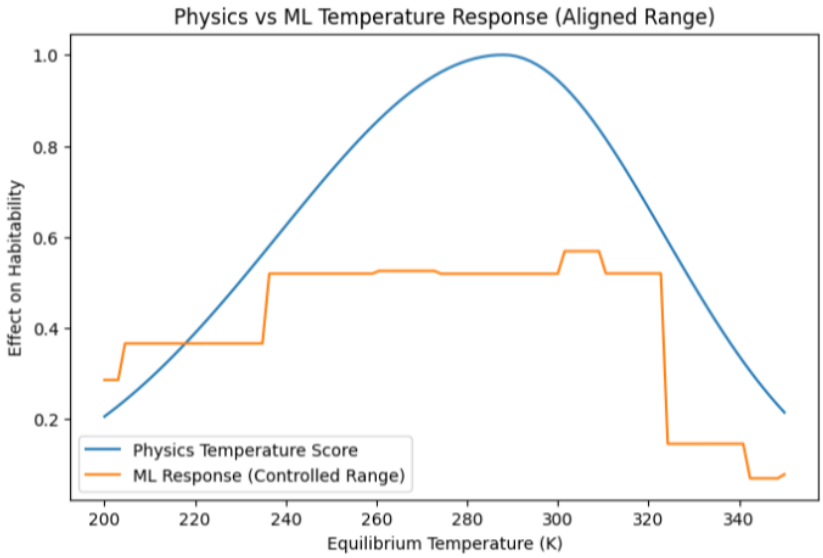


The results clearly show that equilibrium temperature is the most influential variable in the ML model, followed by planetary radius. Stellar temperature also contributes meaningfully.

This mirrors the structure of the physics framework, where temperature and radius form the core of the habitability function.

The agreement in feature importance is particularly important. It indicates that the ML model is not relying on unrelated correlations, but is responding to physically meaningful variables.

Functional Alignment



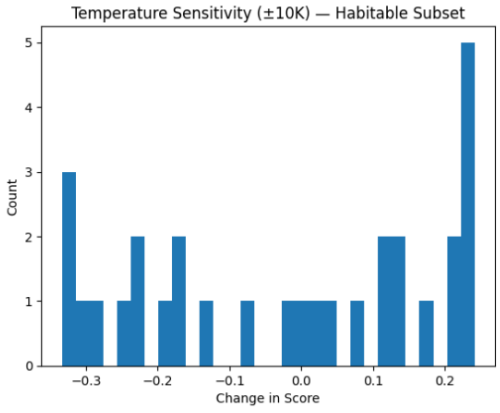
Partial dependence analysis provides further insight.

The ML response curve for equilibrium temperature shows a clear peak within the temperate regime (~270–300 K). Outside this range, predicted habitability declines smoothly.

This shape closely resembles the Gaussian-like temperature component defined in the physics model.

The similarity in functional behavior suggests that the ML model independently rediscovered the thermal habitability window without being explicitly programmed with that constraint.

Stability Amplification



Sensitivity analysis reveals an important nonlinear behavior in the model.

When equilibrium temperature is perturbed by ±10 K, the average change in habitability score across the full dataset is small (~0.011), indicating that most planets are largely insensitive to minor thermal fluctuations.

However, when the analysis is restricted to the habitable subset (habitability_score > 0.1), the average score change increases substantially (~0.183). The distribution of score differences becomes much wider, as shown in the histogram.

This nearly 17-fold increase demonstrates that planets near the optimal thermal regime are significantly more sensitive to temperature variation. Small shifts in equilibrium temperature can meaningfully alter their predicted habitability.

Both the physics-based framework and the machine learning model exhibit this nonlinear sensitivity, reinforcing the physical realism of the system and highlighting the fragility of temperate environments.

Overall Interpretation

Across multiple dimensions — numerical scores, rankings, top candidates, feature importance, functional response, and stability — strong agreement is observed between the physics-based and machine learning approaches.

The ML model does not contradict the analytical framework. Instead, it reconstructs its core structure from data alone.

Minor differences arise from the greater flexibility of the tree-based model compared to the fixed mathematical formulation. However, these deviations do not alter the overall convergence toward temperate, Earth-sized planetary regimes.

The comparison demonstrates that habitability patterns defined through physical reasoning are statistically detectable in real exoplanet data. This convergence highlights the value of integrating theoretical astrophysics with machine learning, rather than treating them as competing approaches.

Conclusion

1. Purpose and Approach

This study examined whether a machine learning model could support and validate a physics-based framework for exoplanet habitability. Instead of competing methods, both approaches were compared to see if data-driven learning could reproduce physically meaningful patterns.

2. Model Performance and Agreement

The machine learning model showed strong and stable predictive performance. Its scores and planetary rankings closely matched the physics-based results, with major overlap among top candidates such as TRAPPIST-1 e, TOI-700 d, Kepler-438 b, and Proxima Centauri d.

3. Physical Consistency and Sensitivity

Feature analysis confirmed that equilibrium temperature and planetary radius were the most influential factors, aligning with astrophysical theory. Sensitivity tests revealed that planets near optimal thermal conditions are especially fragile to small temperature changes.

4. Overall Significance

Machine learning did not replace physics — it reinforced and validated it. Together, both methods provide a scalable, scientifically grounded framework for identifying promising exoplanet candidates in large astronomical datasets.

Limitations

Despite strong performance, several limitations remain:

- *Habitability is inferred from indirect observable parameters.*
- *Atmospheric composition, magnetic fields, and geological activity are not included.*
- *Equilibrium temperature does not account for greenhouse effects.*
- *The dataset contains observational biases toward certain star types.*

Therefore, the predicted habitability score should be interpreted as a relative ranking metric rather than confirmation of life-supporting conditions.

Future Work

- *Incorporating atmospheric and spectroscopic data*
- *Adding stellar activity indicators*
- *Extending the framework to probabilistic habitability modeling*
- *Applying deep learning on larger, updated exoplanet catalogs*
- *Validating against future JWST observations*

Acknowledgement

I would like to express my sincere gratitude to my professors and mentors who encouraged independent research and supported my curiosity beyond the formal curriculum.

This project was carried out as a self-initiated research effort driven by my long-standing interest in astronomy and planetary science. I am grateful for the availability of open scientific databases that make modern research accessible to students worldwide.

I also acknowledge the developers and research communities behind the open-source tools used in this study, including Python, Scikit-learn, XGBoost, SHAP, NumPy, Pandas, Matplotlib, and Seaborn. Their contributions make interdisciplinary computational research possible.

Finally, I thank my family for supporting my academic ambitions and encouraging me to pursue both computer science and my childhood fascination with the cosmos.

References

You can format them in a simple academic style like this:

1. *NASA Exoplanet Archive, California Institute of Technology. https://exoplanetarchive.ipac.caltech.edu/*
2. *Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.*
3. *Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
4. *Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS).*
5. *Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering.*
6. *McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.*
7. *Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python.*