Problem Statement

The objective of this project is to develop a Machine Learning model that classifies exoplanets into habitable, marginally habitable, and non-habitable categories based on planetary and stellar properties. The goal is to determine which exoplanets are most likely to support life using a scientifically derived Habitability Index.

Assumptions

- Exoplanets with stronger gravity can retain an atmosphere, which is critical for life.
- Stellar radiation plays a major role in planetary habitability.
- Earth Similarity Index (ESI) is a reliable metric for comparing exoplanets to Earth.
- P_Eccentricity**(Orbital Shape):**Planets with highly **eccentric orbits** experience **extreme temperature variations**, affecting climate stability.A stable **low-eccentricity orbit** is preferable for habitability.
- **P_TEMP_EQUIL (Equilibrium Temperature):**Helps estimate **surface temperature** based on distance from the star.Critical for identifying planets within the **habitable zone** where liquid water can exist.
- **S_TEMPERATURE (Star Temperature):**Affects **stellar radiation received** by planet.Determines whether a planet is **too hot or cold** for life.

The dataset provides sufficient features to predict habitability accurately.

Custom Habitability Index: The Habitability Index (HI) is a weighted metric based on key planetary properties:

Formula:

HI = (0.4 * ESI) + (0.3 * Atmospheric Retention) + (0.3 * Stellar Radiation)

Feature Weights Explanation: -

Earth Similarity Index (40%) - Measures similarity to Earth in size and radiation. - Atmospheric Retention (30%) - Higher gravity helps retain an atmosphere. - Stellar Radiation (30%) - Too much radiation can be detrimental to life.

DATA PREPROCESSING

- Handled missing values through imputation.
- Encoding categorical data to numerical and vice versa
- Removed highly correlated redundant features.
- Identified and corrected outliers using IQR and log transformation.Using the help of boxplot.
- Rechecked outliers using heatmap.

FEATURE ENGINEERING

- Created the Habitability Index (HI) using ESI, Gravity, and Stellar Radiation.
- Selected key features for model training.

MODEL TRAINING

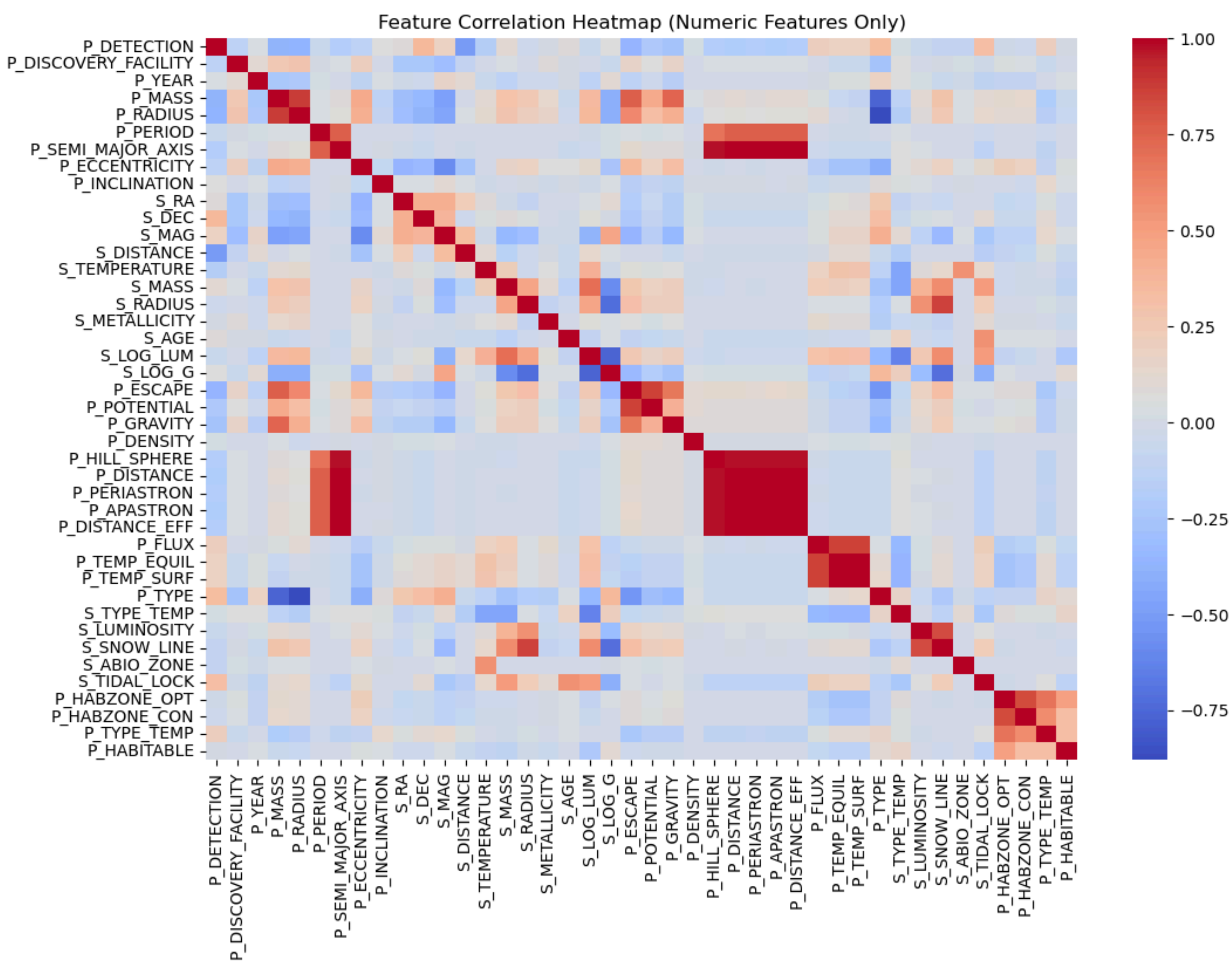
- Used Random Forest and attempted XGBoost (but faced installation issues).
- Applied Standard Scaling to normalize features.
- Trained and evaluated classification models to predict habitability.

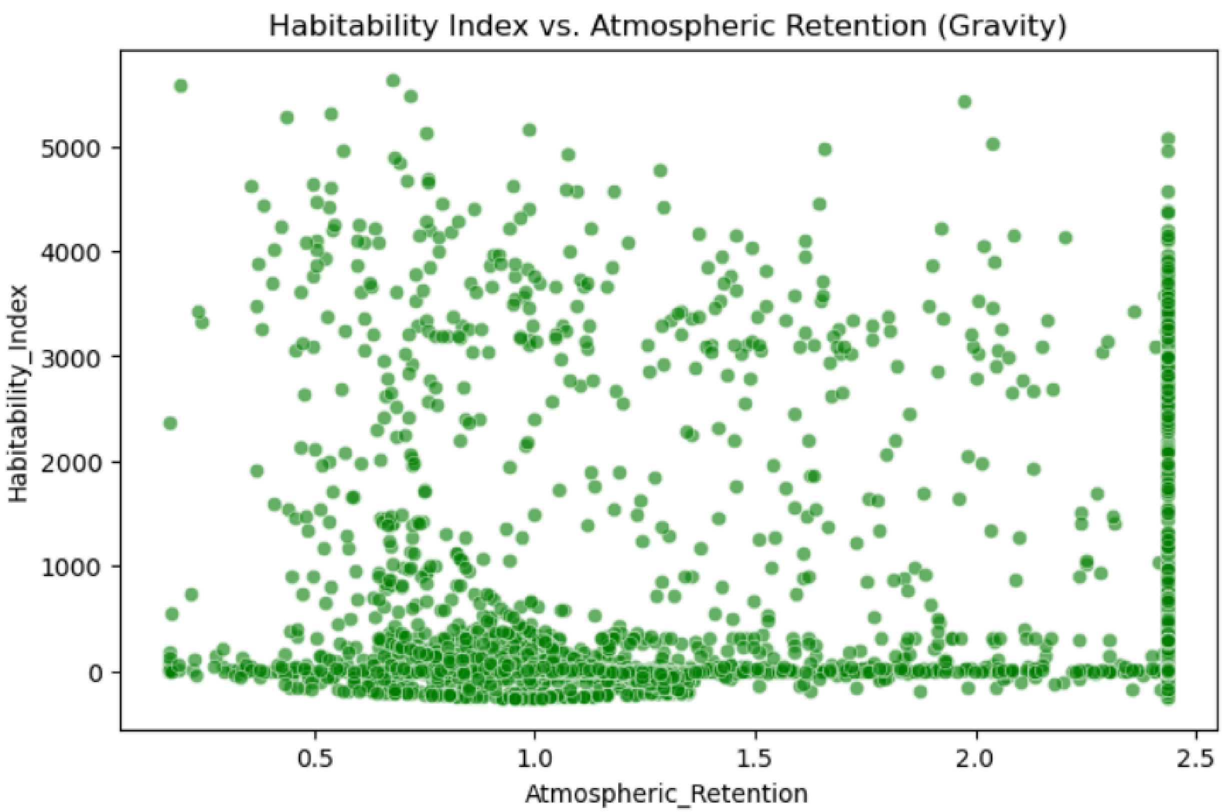
Findings & Recommendations: 1.

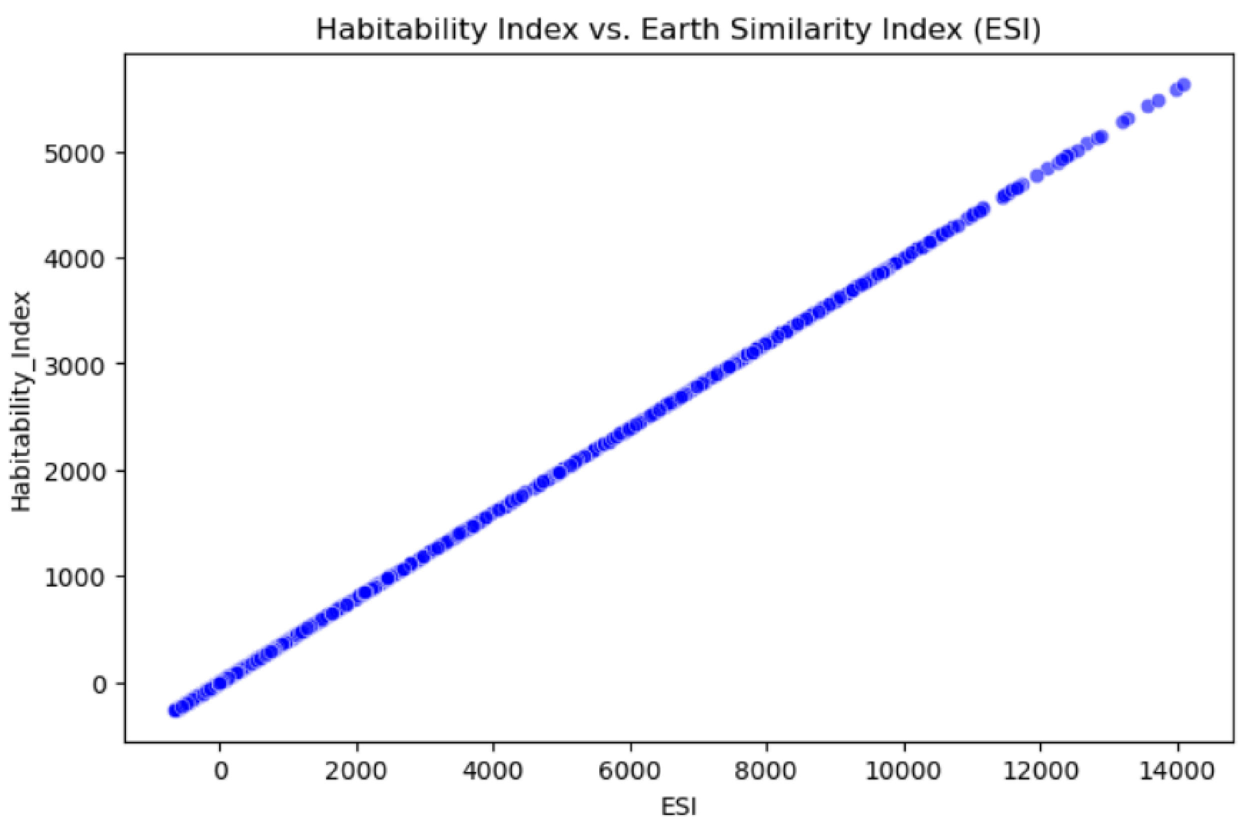
1.The Random Forest model achieved 99.55% accuracy, with perfect classification for non-habitable planets.
2. However, the model struggled with classifying habitable planets due to class imbalance.
3. XGBoost was not available, but LightGBM or SMOTE (oversampling) could further improve results.
4. Future enhancements should explore additional planetary features like atmospheric composition and water presence.

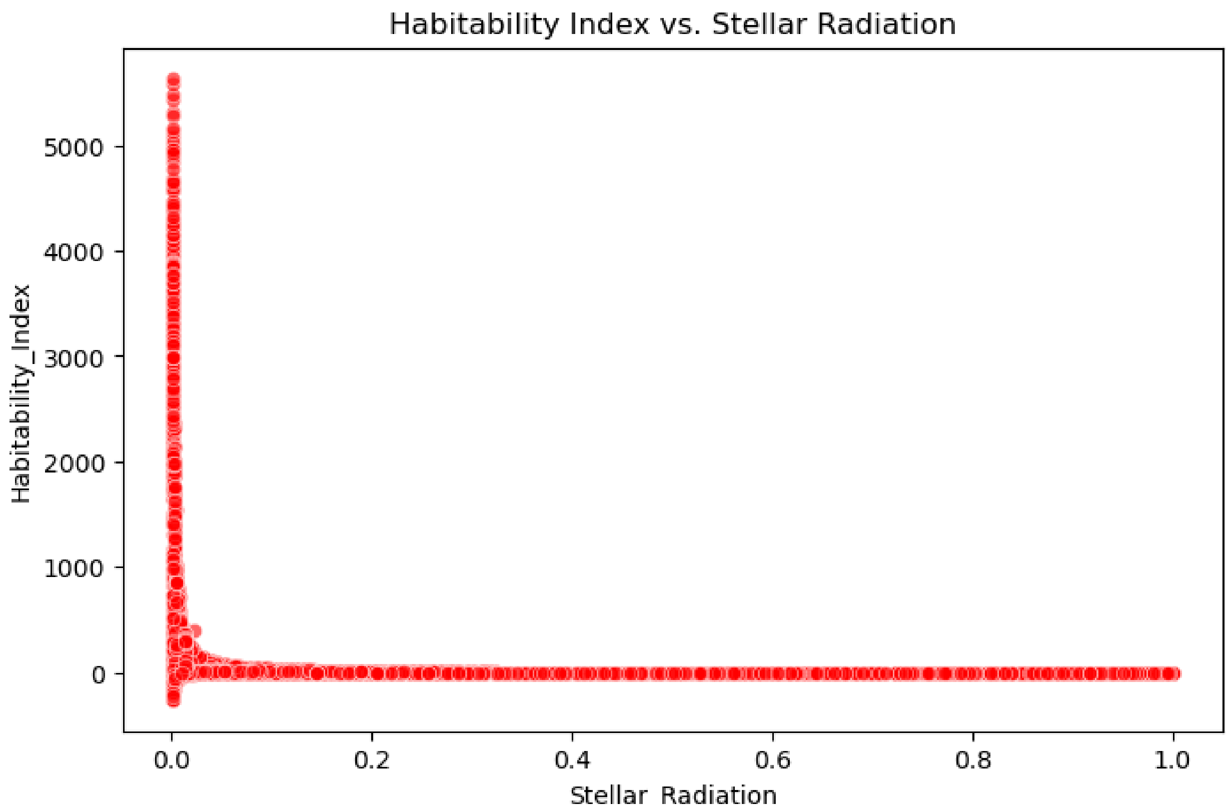
Appendix: Charts & References

- Correlation heatmap confirmed redundant features were removed.




Habitability Index vs. Atmospheric Retention (Gravity)
Habitability_Index
Atmospheric_Retention
0
1000
2000
3000
4000
5000
0.5
1.0
1.5
2.0
2.5


Habitability Index vs. Earth Similarity Index (ESI)
Habitability_Index
ESI
0
1000
2000
3000
4000
5000
0
2000
4000
6000
8000
10000
12000
14000



- Scatter plots showed logical relationships between features and habitability.
- Feature importance analysis proved that HI, ESI, and gravity are key determinants.
- Dataset: Exoplanet Dataset

https://nbviewer.org/github/AISH2211Byte/STELLER-ANALYTICS-JYUPYTER-NOTEBOOK/blob/main/STELLERNEW1.ipynb
https://github.com/AISH2211Byte/STELLER-ANALYTICS-JYUPYTER-NOTEBOOK/blob/main/STELLERNEW1.ipynb