Milestone 2

1 Introduction

- Project objective: Predict exoplanet habitability
- Why habitability prediction matters

2 Dataset Description

- Source: NASA Exoplanet Archive
- Number of rows: 39,212 (example from your CSV)
- Number of features: 9 (numeric features you preprocessed)

3 Data Preprocessing

- Handling missing values: Filled with median
- Feature selection: Only numeric features used for ML
- Feature scaling: StandardScaler applied
- Why preprocessing is important

4 Machine Learning Model

- Model chosen: **Random Forest Regressor** (continuous target)
- Reason for choice: Handles regression, non-linear relationships
- Training process: 80/20 train-test split, trained on all features

5 Model Evaluation

- Metrics used: Mean Squared Error (MSE): 0.19271975164337254, R^2 Score: 0.8076497073829525
- Interpretation: MSE = 0.19, $R^2 = 0.81 \rightarrow$ good prediction accuracy

6 Conclusion

- Final observations: Model trained successfully, ML-ready CSV generated
- Future improvements: Feature engineering, hyperparameter tuning, try other models