TITLE:

Exoplanet Habitability Prediction System

Name: Rushitha Konangi
Branch: Electronics and Communication Engineering

MILESTONE 1:

1. Introduction

The rapid growth of astronomical data from missions such as NASA Kepler and TESS has enabled large-scale discovery of exoplanets. This project applies machine learning techniques to predict the habitability of exoplanets using planetary and stellar features.

2. Problem Statement

The project aims to classify exoplanets as habitable or non-habitable. Challenges include missing values, outliers, absence of direct habitability labels, and severe class imbalance.

3. Dataset Description

The dataset is sourced from the NASA Exoplanet Archive via Kaggle. It includes planet radius, mass, orbital distance, equilibrium temperature, and host star properties.

4. Data Preprocessing

Data preprocessing involved removing irrelevant columns, handling missing values using mean/median imputation, and treating outliers using statistical techniques.

5. Exploratory Data Analysis

EDA was performed using histograms, boxplots, scatter plots, and correlation heatmaps to understand feature distributions and relationships affecting habitability.

During the first milestone, the focus was on data collection and initial preparation. Exoplanet datasets were successfully collected from reliable sources such as the NASA Exoplanet Archive and Kaggle. The collected data was properly organized and stored in a structured CSV format with a well-defined schema. Basic data cleaning techniques were applied, including handling missing values and removing irrelevant attributes. Initial feature engineering was performed to generate habitability-related metrics. Additionally, an initial data dictionary was documented, explaining each feature and its relevance to exoplanet habitability prediction.

MILESTONE 2:

1. System Purpose

The Exoplanet Habitability Prediction System is an end-to-end machine learning application designed to classify and rank exoplanets based on their potential to support life. The system processes astronomical datasets, applies statistical preprocessing and feature engineering, trains supervised learning models, and exposes predictions through RESTful APIs.

2. High-Level Architecture

Pipeline Architecture:

Raw Dataset
↓
Data Cleaning & Validation
↓
Feature Engineering
↓
Scaling & Dimensionality Reduction
↓
ML Model Training
↓
Model Evaluation
↓
REST API Inference Layer
↓
Visualization & Ranking

In the second milestone, the dataset was prepared for machine learning model development. The cleaned dataset was split into training and testing sets using an 80:20 ratio. Machine learning pipelines were implemented to handle feature scaling and encoding. Models such as Random Forest and XGBoost were trained and evaluated using suitable performance metrics. Initial habitability prediction accuracy was analyzed, different models were compared, and exoplanets were ranked based on predicted habitability.

MILESTONE 3:

Technology Stack

Backend & ML: Python 3.x, FastAPI (ASGI), NumPy, Pandas, Scikit-learn, XGBoost
Visualization: Matplotlib, Seaborn, Chart.js
Deployment & Runtime: Render Cloud, Uvicorn ASGI server, GitHub for CI/CD

Dataset Engineering

Dataset Source: NASA Exoplanet Archive (Kaggle mirror)
Features:

- Planetary: Radius, Mass, Density, Orbital distance, Equilibrium temperature
- Stellar: Stellar temperature, Luminosity, Radiation characteristics
  Target Variable: Binary classification (1 → Habitable, 0 → Non-Habitable)

Data Preprocessing Pipeline

- Missing value handling (mean/median imputation)
- Columns with excessive missingness removed
- Outlier treatment using Z-score and IQR
- Data validation with range checks and statistical verification

Feature Engineering

- Habitability Score Index (weighted combination of radius, temperature, orbital distance)
- Stellar Compatibility Index (derived from stellar properties)
- Scaling with StandardScaler
- Dimensionality reduction using PCA

Class Imbalance Strategy

- Stratified train-test split (80:20)
- SMOTE applied only on training data to avoid bias toward majority class

Machine Learning Models

- Logistic Regression: Baseline linear classifier
- Support Vector Machine (SVM): Kernel-based (RBF) for non-linear separations

- Random Forest: Ensemble decision trees (final model with best performance)
- XGBoost: Gradient boosting for multi-class habitability prediction

Model Training & Evaluation

- Training/Testing split: 80/20
- Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC
- Validation: Confusion matrix, ROC curve, feature importance ranking

Prediction & Ranking Logic

- Model outputs probability scores
- Exoplanets ranked based on predicted habitability
- Top-N most habitable exoplanets identified

The third milestone focused on backend development and system integration. A RESTful backend API was developed to integrate the trained machine learning models. API endpoints were implemented to accept exoplanet input parameters and return habitability predictions. The API responses were structured in JSON format. Initial frontend integration was completed to allow users to interact with the prediction system and view results.

MILESTONE 4:

Visualization Engine

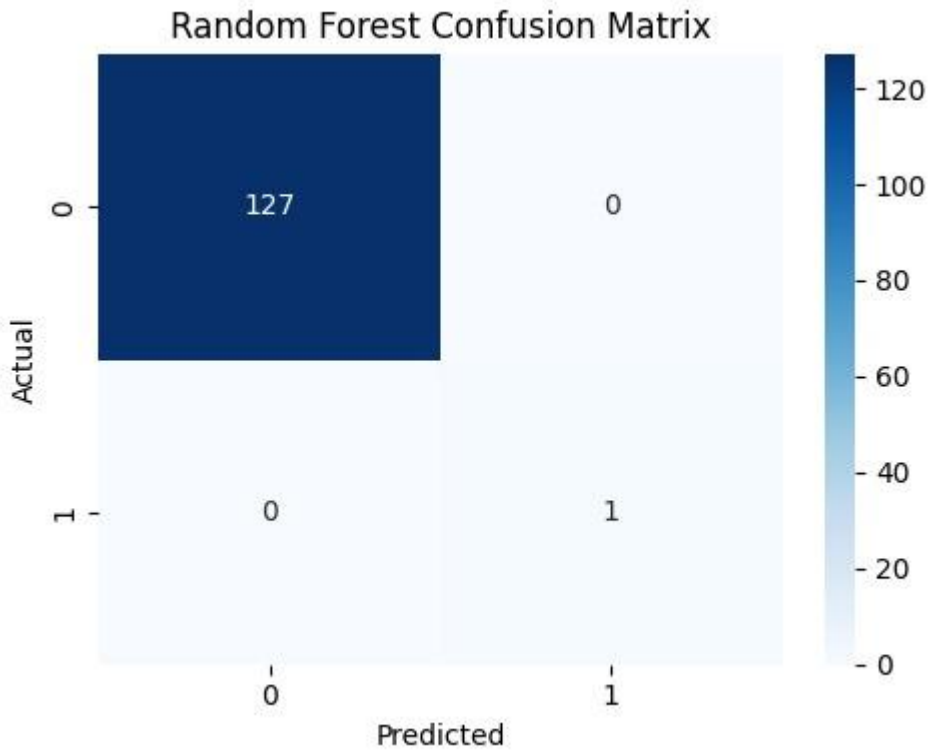
- Feature importance bar charts
- Habitability score distributions
- Scatter plots (orbital distance vs radius)
- Correlation heatmaps

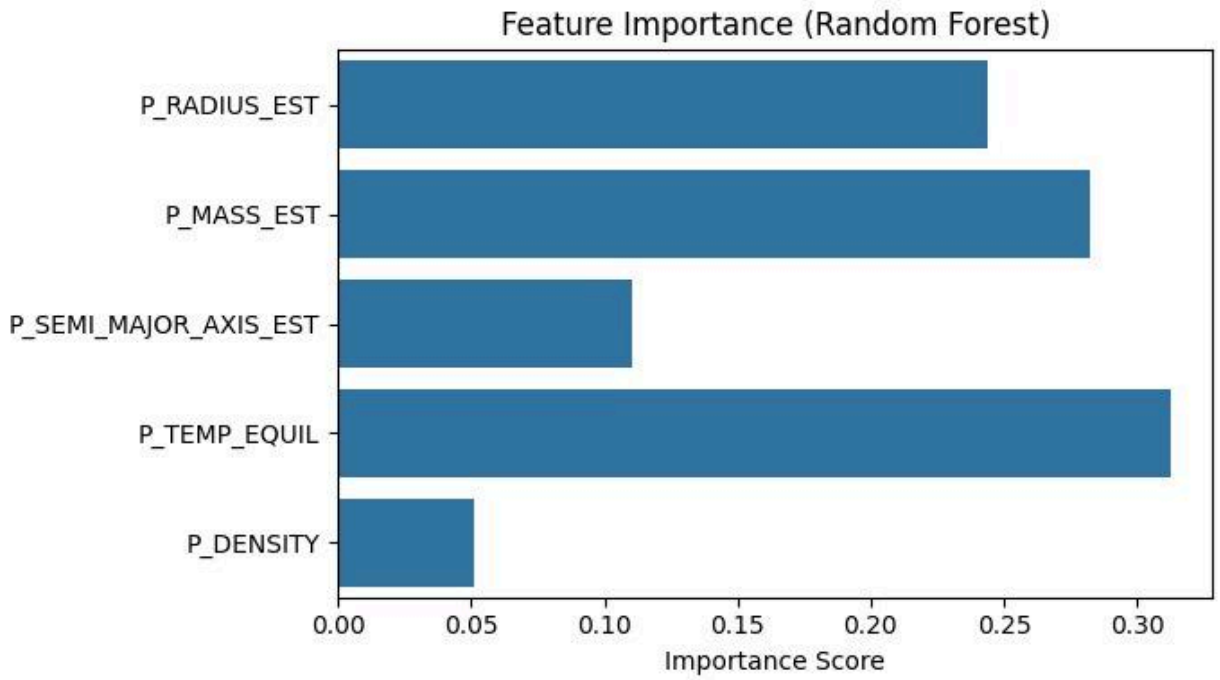
System Limitations

- Sparse and imbalanced astronomical datasets
- Predictions depend on indirect habitability indicators
- Potential overfitting in ensemble models

Future Enhancements

- CNN models for light-curve analysis
- SHAP-based explainable AI
- Real-time data integration
- Advanced dashboard interactivity

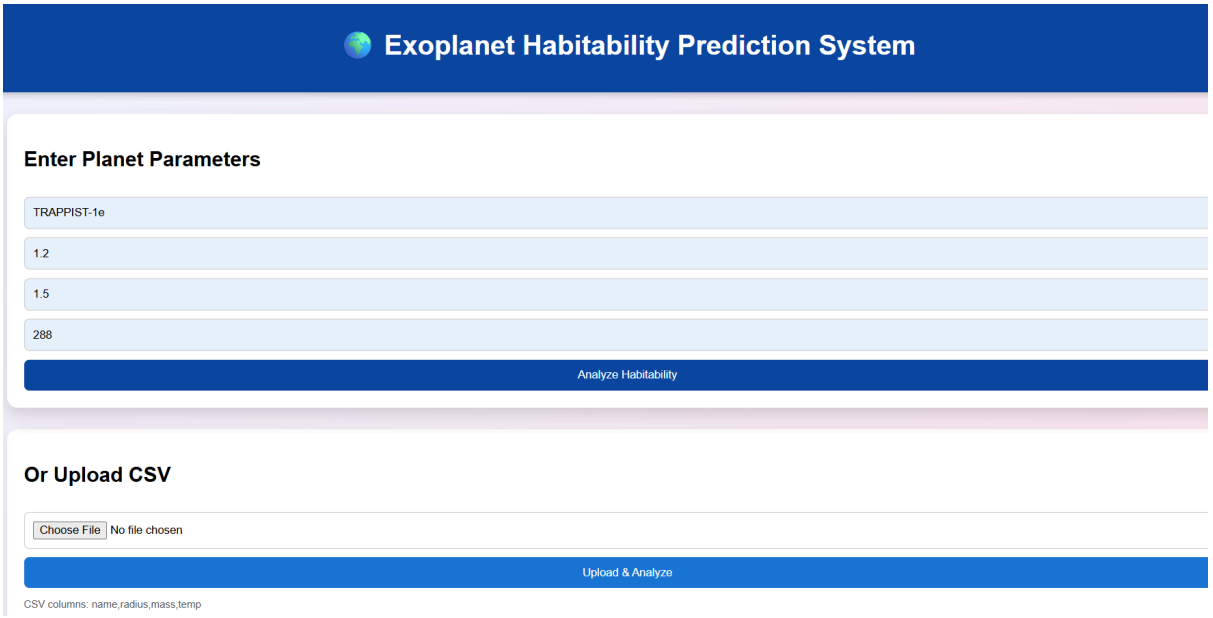


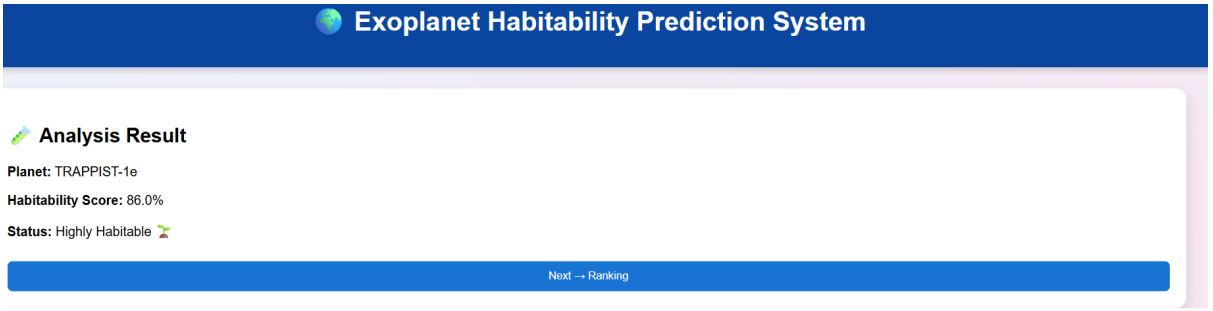


The final milestone emphasized visualization, deployment, and documentation. An interactive dashboard was developed to visualize habitability insights, feature importance, and prediction results. The system supports exporting habitability reports and insights in formats such as PDF and Excel. Full integration of the frontend, backend, and machine learning model was achieved. The complete application was successfully deployed on a cloud platform, and comprehensive technical documentation was prepared.

RESULTS

FRONTEND

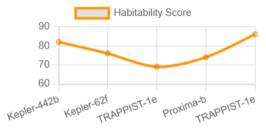




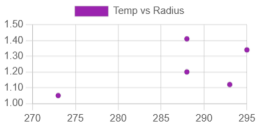
🌍 Exoplanet Habitability Prediction System

📊 Visualization Dashboard

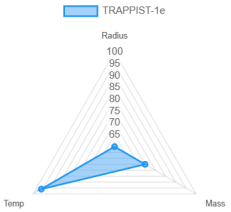
Score Distribution



Temperature vs Radius

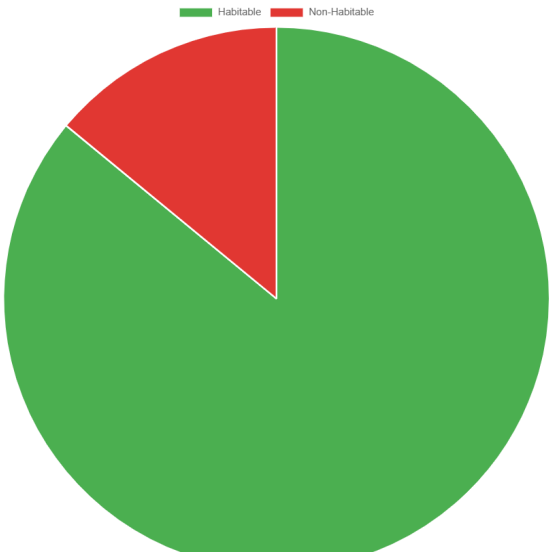


Planet Profile (Radar)

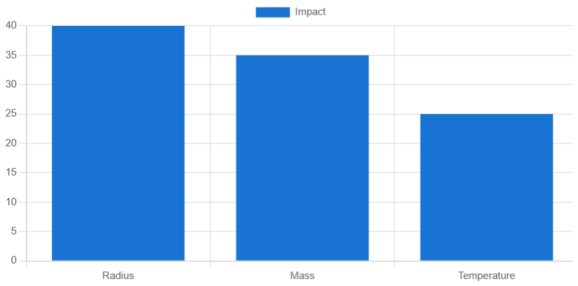


Top Planets by Score

Habitability Pie Chart



Feature Contribution



CONCULSION

This project successfully demonstrates the application of machine learning techniques for predicting the habitability of exoplanets using planetary and stellar parameters. By leveraging data from the NASA Exoplanet Archive, the system effectively handled real-world challenges such as missing values, outliers, and severe class imbalance.

Through systematic data preprocessing, feature engineering, and exploratory data analysis, meaningful habitability-related features were extracted. Advanced techniques such as SMOTE and PCA improved model robustness and learning efficiency. Multiple machine learning models were implemented and evaluated, among which the Random Forest classifier achieved the best overall performance in predicting exoplanet habitability.

The project further extended beyond model development by integrating a RESTful backend API, an interactive frontend dashboard, and cloud deployment. This end-to-end implementation enables real-time habitability prediction, visualization, and ranking of exoplanets, making the system practical and scalable for real-world use.

Overall, this project highlights the effectiveness of machine learning in astronomical data analysis and provides a reliable, extensible framework for assisting researchers in identifying potentially habitable exoplanets for further scientific exploration.

REFERENCES

Video Demo:

https://onedrive.live.com/?qt=allmyphotos&photosData=%2Fshare%2F7455FACDCC191830%21s7ba3f36466c141838e8686b421bd15d1%3Fithint%3Dvideo%26e%3DGx7JbM%26migratedtospo%3Dtrue&cid=7455FACDCC191830&id=7455FACDCC191830%21s7ba3f36466c141838e8686b421bd15d1&redeem=aHR0cHM6Ly8xZHJ2Lm1zL3YvYy83NDU1ZmFjZGNjMTkxODMwL0lRQms4Nk43d1dhRFFZNkdoclFodlJYUkFaUUhGSDJZb3ItSVAxand4RzQzQ1dnP2U9R3g3SmJN&v=photos

Live Application URL:

https://habitability-of-exoplanets-2.onrender.com/

GitHubRepository:

https://github.com/springboardmentor74280b-design/Habitability-of-Exoplanets/tree/rushitha-konangi