Project Documentation

Predicting the Habitability of Exoplanets Using Machine Learning

1. Introduction

In recent years, scientists have discovered thousands of planets outside our solar system, called exoplanets. However, not all of these planets are suitable for supporting life. Determining which planets may be habitable is difficult because many factors must be analyzed, such as temperature, size, distance from the star, and atmospheric conditions.

The goal of this project is to use ***Machine Learning techniques*** *to analyze exoplanet data and predict whether a planet may be habitable or not. This system helps researchers and students understand how data science can be applied in astronomy.*

---

2. Objective of the Project

The main objectives of this project are:

- *To collect and analyze exoplanet data*
- *To preprocess and clean the dataset*
- *To train a machine learning model for prediction*
- *To display predictions in a simple interface*
- *To visualize important factors affecting habitability*

---

3. Tools and Technologies Used

Programming Language:

- Python

Libraries:

- Pandas
- NumPy
- Scikit-learn
- Matplotlib

Web Technologies:

- HTML
- CSS
- JavaScript

Framework:

- Flask

Development Tools:

- VS Code
- GitHub

---

4. Methodology

The project is divided into several steps:

Step 1: Data Collection

The dataset is collected from publicly available sources such as NASA or Kaggle. The dataset contains information about planets and stars.

Step 2: Data Cleaning

The collected data may contain missing values or unnecessary columns. These issues are handled by:

- Removing null values
- Normalizing numerical features
- Encoding categorical data

Step 3: Feature Selection

Important parameters like radius, mass, temperature, and orbital period are selected for training the model.

Step 4: Model Training

Machine learning algorithms such as:

- Random Forest
- Logistic Regression

are used to train the prediction model.

Step 5: Prediction

The trained model predicts whether a planet is habitable based on input parameters.

Step 6: Visualization

Graphs and charts are used to understand the data and prediction results.

---

5. System Architecture

The system consists of three main parts:

1. Dataset and preprocessing
2. Machine learning model

3. User interface and visualization

All components work together to provide predictions.

---

6. Expected Results

The system should be able to:

- Predict habitability of planets
- Show results clearly
- Help in understanding important factors affecting habitability

---

7. Applications of the Project

This project can be useful in:

- Space research
- Data science learning
- Astronomy education
- Machine learning practice

---

8. Advantages of the System

The proposed system provides several advantages:

- Machine learning helps analyze large astronomical datasets efficiently.
- The prediction process is faster compared to manual analysis.
- Visualization helps in better understanding of planetary characteristics.

- The system can be improved in the future by adding more datasets and advanced models.

---

9. Limitations of the System

Every project has some limitations. In this system:

- The prediction depends on the quality of the dataset.
- Some planetary parameters may be missing or incomplete.
- The model predictions are probabilistic and not fully accurate.

---

10. Future Enhancements

This project can be improved further by:

- Using deep learning models for better prediction accuracy.
- Integrating real-time NASA datasets.
- Improving the user interface and dashboard.
- Deploying the application on a cloud platform for public access.

Conclusion

This project demonstrates how machine learning can be applied to real-world scientific problems. By analyzing planetary data, the system provides predictions that can help in identifying potentially habitable exoplanets.

It also helps students understand the full workflow of a data science project, from data collection to deployment.