ML PROJECT ON

EXOPLANET HABITABILITY PREDICTION

By-

Name- Pranjal Sinha

Reg. no.- 12206214

Roll no.- B53

INTRODUCTION

Exploring Earth-like Worlds Beyond Our Solar System

Thousands of exoplanets have been discovered, but finding those like Earth remains a major goal.

Instead of traditional observation methods, this project uses unsupervised machine learning to group exoplanets by Earth-likeness — uncovering patterns without needing labeled data.

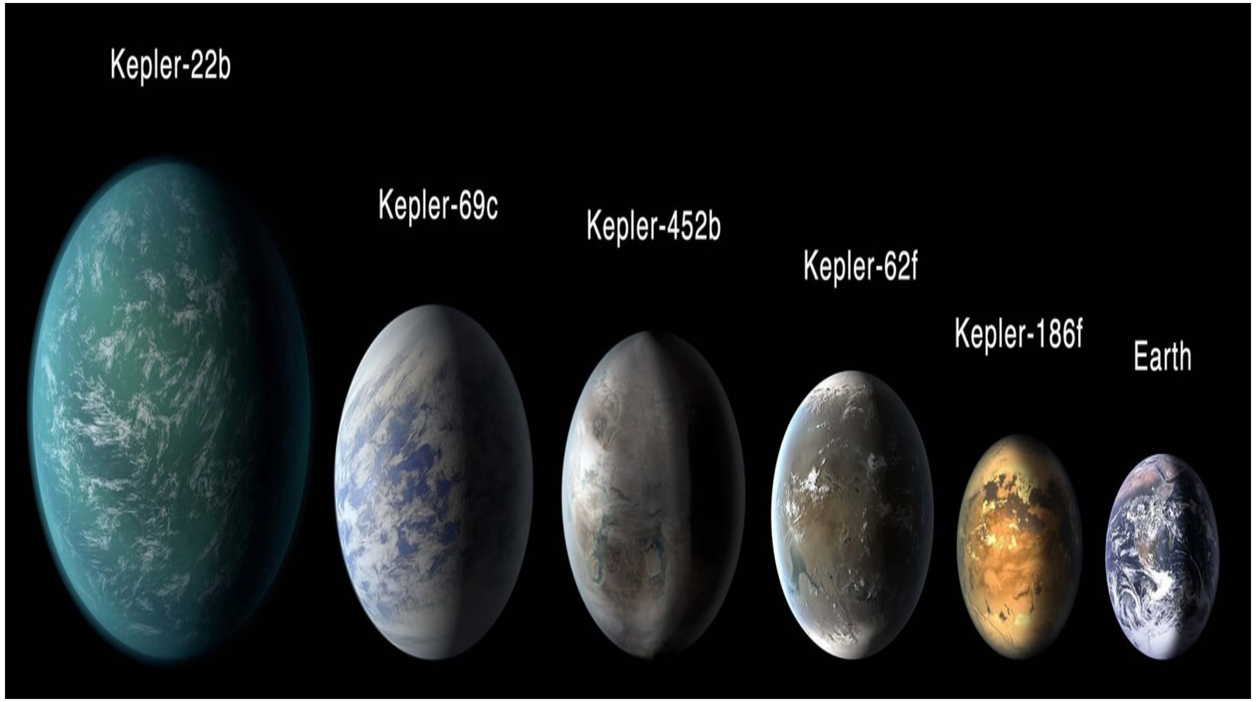
PROBLEM STATEMENT

From Habitability Classification to Earth-Likeness Clustering

The original goal was to classify exoplanets based on their habitability. However, due to the lack of labeled data indicating which exoplanets are actually habitable, we shifted to an unsupervised learning approach.

Our Goals:

- Group similar exoplanets using unsupervised learning
- Measure Earth-likeness by comparing to Earth's parameters
- Build a webpage for users to input planet data and get insights



DATASET DESCRIPTION

Description:

- 5000+ confirmed exoplanets.
- Over 90 features including planetary and stellar properties.
- Key parameters include orbital period, radius, mass, equilibrium temperature, stellar temperature, and more.

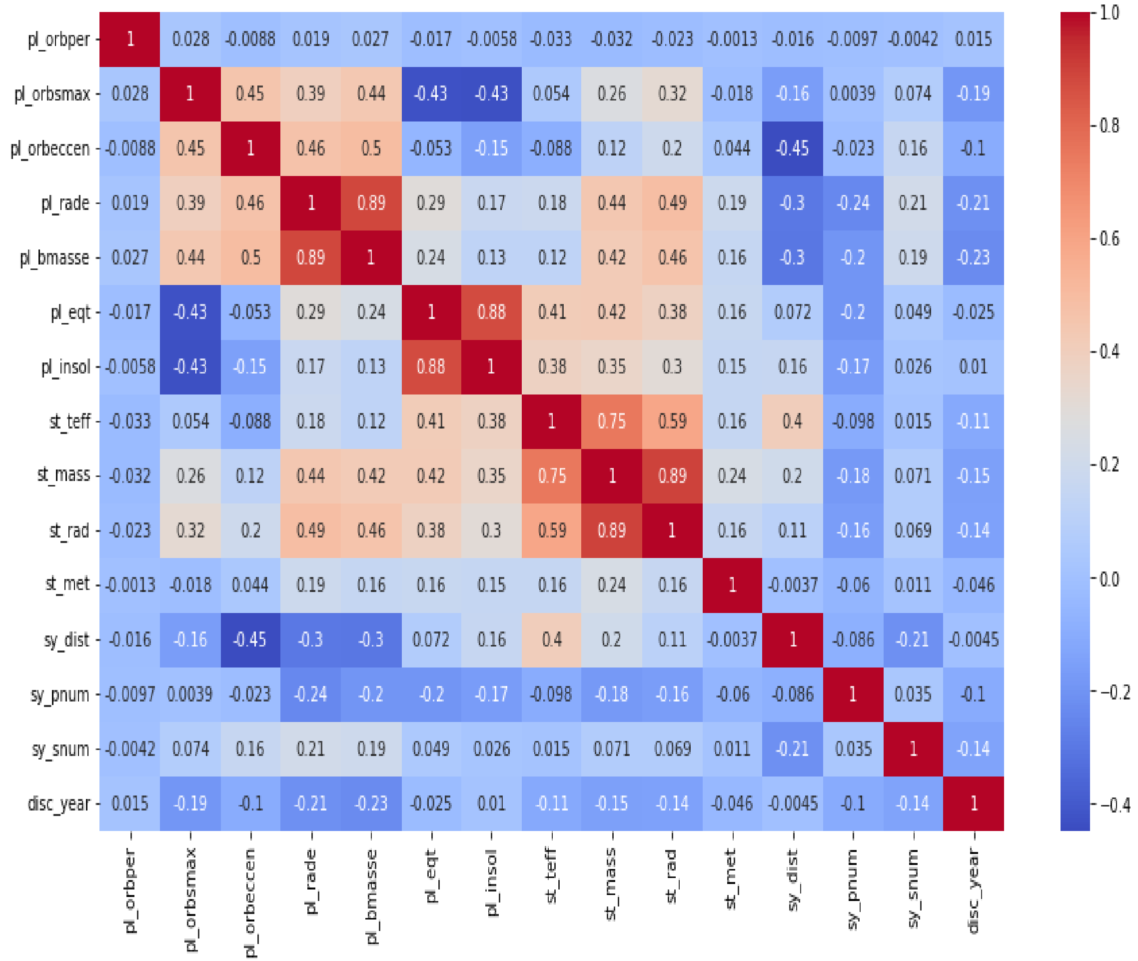
Data Cleaning Involved:

- Handling missing values.
- Converting units to Earth-relative scales.
- Filtering entries with critical missing data.

KEY FEATURES USED

A variety of features were used after careful EDA and domain research:-

1) pl_orbper: Orbital Period
2) pl_rade: Planet Radius (Earth units)
3) pl_bmasse: Planet Mass (Earth units)
4) pl_eqt: Equilibrium Temperature
5) st_teff: Stellar Effective Temperature
6) st_rad: Stellar Radius
7) st_mass: Stellar Mass
8) pl_insol: Insolation Flux
9) pl_orbeccen: Orbital Eccentricity

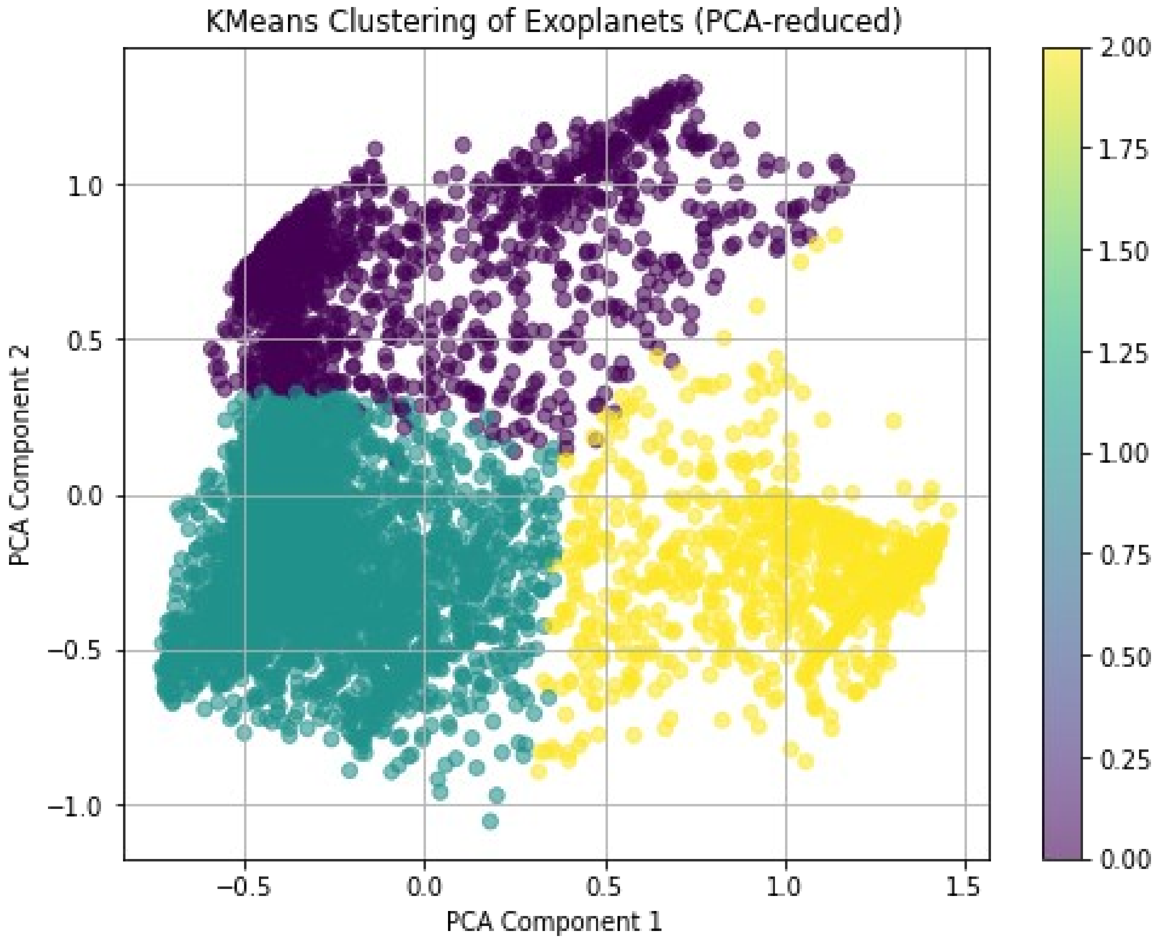


These features directly influence planetary conditions and potential for habitability.

FEATURE ENGINEERING

Data Preparation at a Glance

- **Missing Values:** Dropped critical nulls, filled the rest with median/mode
- **Scaling:** Used StandardScaler to normalize all features
- **Dimensionality Reduction:** Applied PCA (2D) while retaining ~90% variance
- **Earth Similarity Index:** Evaluated, but skipped to maintain unsupervised integrity



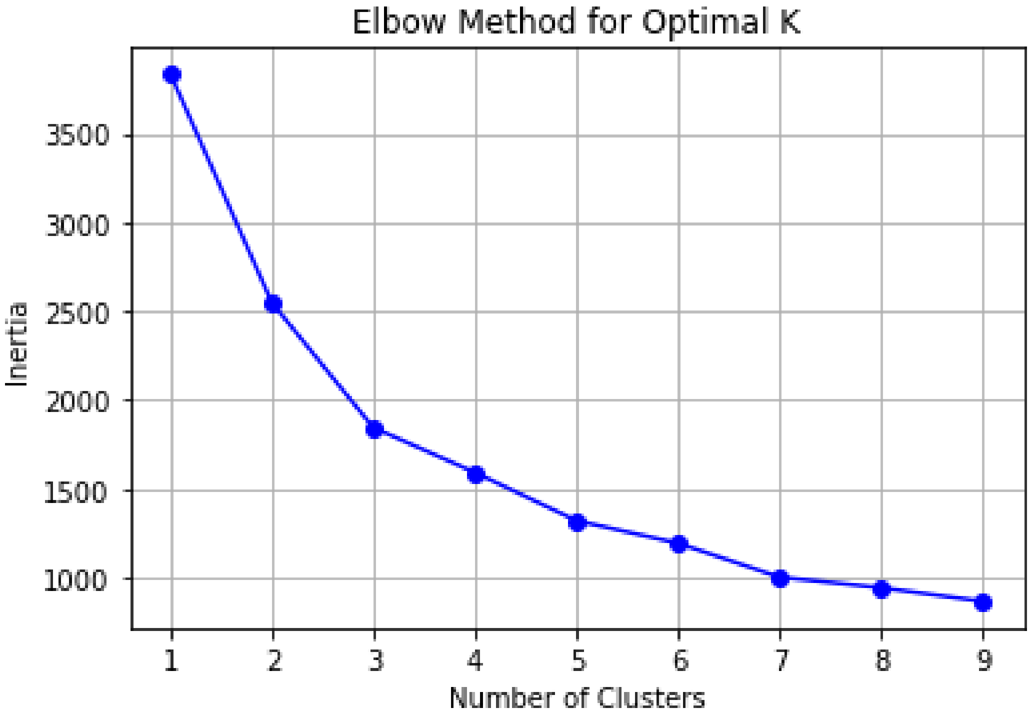
MODEL SELECTION

Clustering Approach

- Tried **KMeans** — simple, fast, and effective
- Chose **K = 3** using the Elbow Method

Final Clusters:

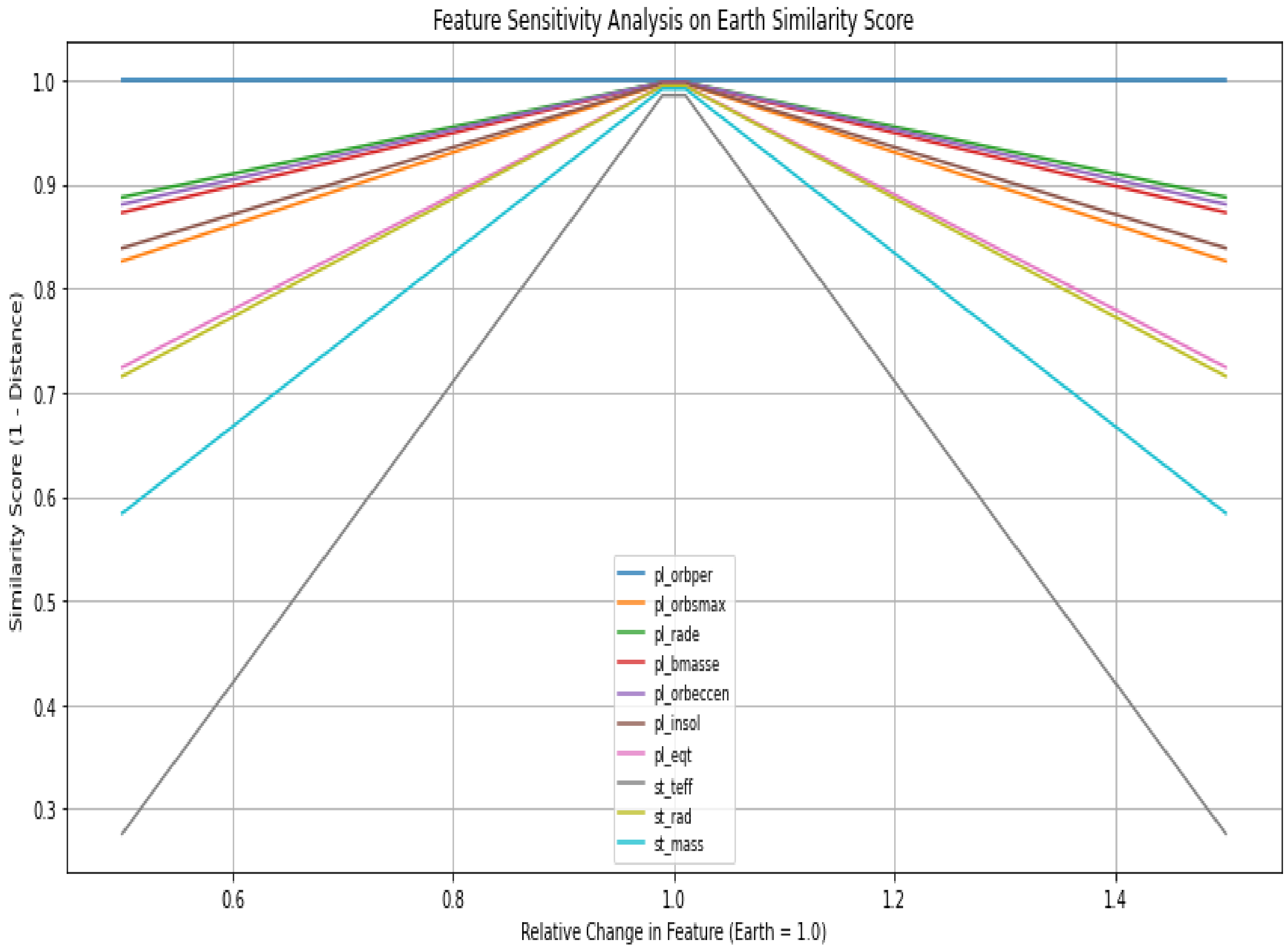
- **Cluster 0:** Earth-like
- **Cluster 1:** Intermediate
- **Cluster 2:** Unlikely to be habitable



RESULTS

KMeans successfully divided the dataset into 3 clusters:

- Cluster 0: Close to Earth's parameters (High Habitability)
- Cluster 1: Moderately habitable planets
- Cluster 2: Hot Jupiters, rogue planets, or extreme orbits (Low Habitability)
- Visualized using PCA-reduced scatter plot. Distribution validated through cluster centers.



WEB APPLICATION INTEGRATION

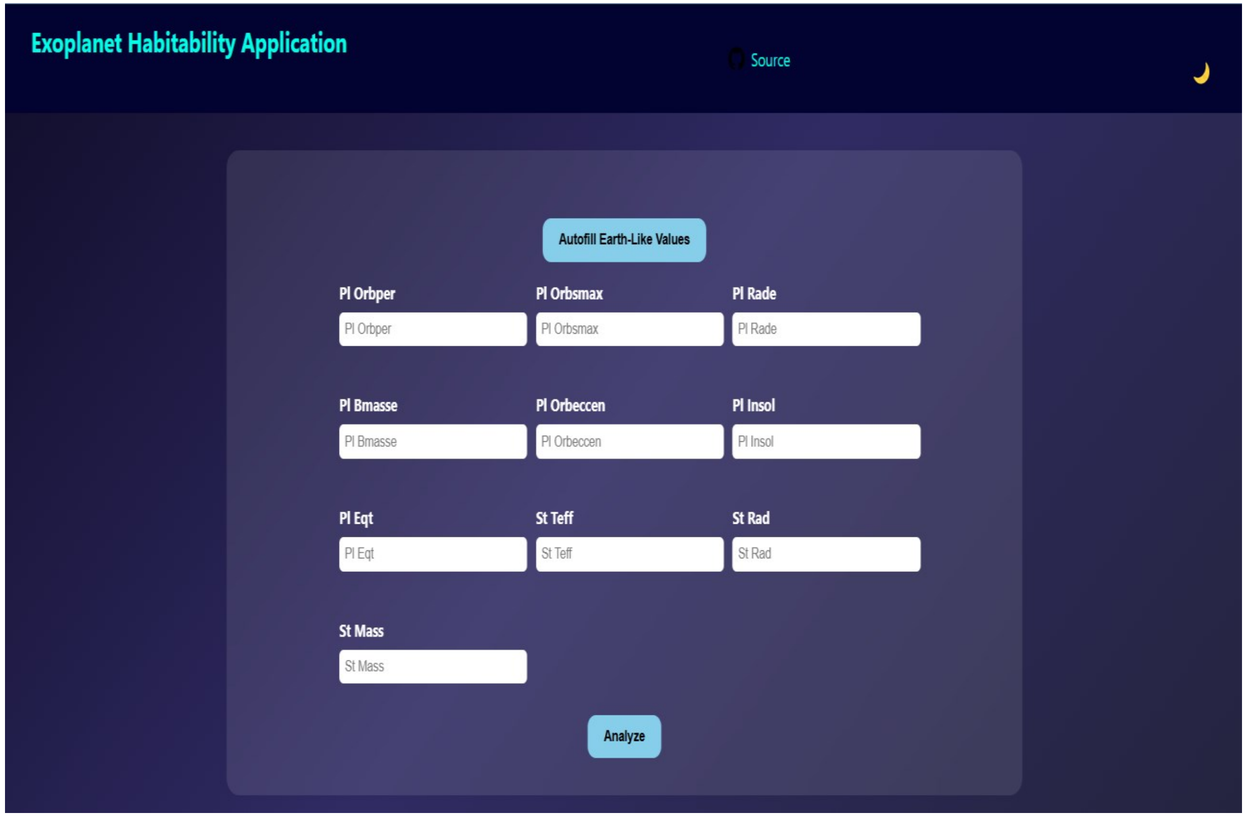
Built with: Python (Flask), HTML/CSS, Bootstrap

What it does:

- Users enter planet data
- Model predicts habitability cluster

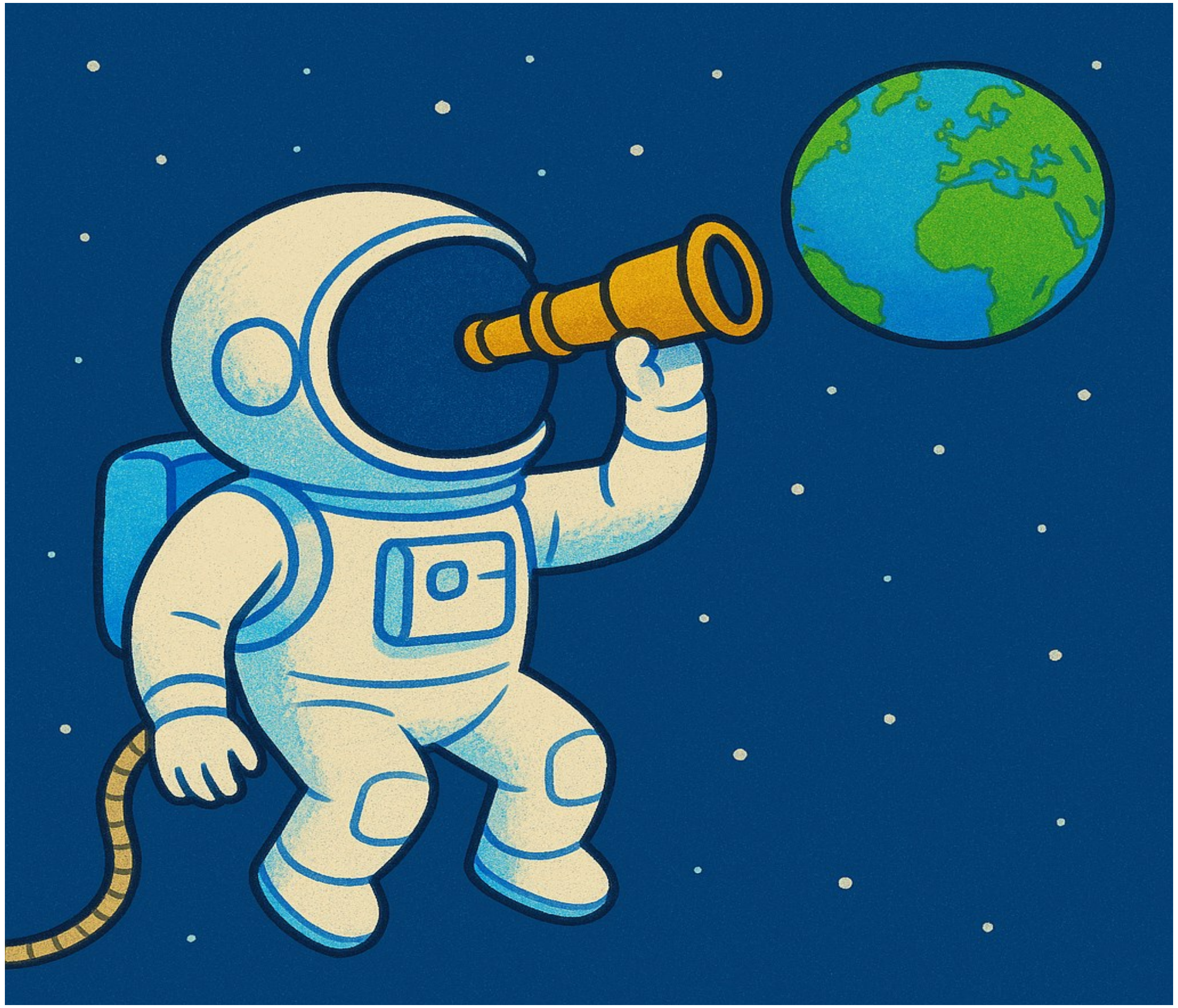
Key Features:

- Input validation
- Real-time results
- Simple, user-friendly design



CHALLENGES FACED

- Missing Data: Several critical columns had over 30% missing values.
- Lack of Labels: Forced us to adopt unsupervised learning.
- Feature Selection: Required astrophysical knowledge.
- Cluster Interpretation: Naming clusters without labels was subjective.
- App Integration: Linking backend ML model with a clean frontend took time.



CONCLUSION

In conclusion, this project successfully demonstrates the use of unsupervised machine learning to assess the habitability of exoplanets based on their similarity to Earth.

By leveraging astrophysical data and clustering techniques, we shifted from a traditional classification approach to a more exploratory analysis that enables insights into planetary characteristics without requiring predefined labels.

The development of a web application further enhances accessibility and practical application of the model. This work not only deepens our understanding of exoplanetary systems but also lays the foundation for future research in astrobiology and automated planetary analysis, with potential for real-world scientific exploration.

REFERENCES

- NASA Exoplanet Archive: https://exoplanetarchive.ipac.caltech.edu/docs/data.html
- Astrobiology papers from journals like AAS, Elsevier.
- Scikit-learn Documentation
- Flask Documentation
- Open source astronomy datasets

Thank You

Thank you for your time and attention! I appreciate the opportunity to share my project. A special thanks to my mentors, peers, and the faculty at Lovely Professional University for their support. I hope this presentation offered valuable insights into the role of machine learning in discovering habitable worlds beyond Earth.

GitHub Repository Link

Questions and feedback are welcomed!