PREDICTING THE HABITABILITY OF EXOPLANETS USING ML

MILE STONE-1 DATASET DESCRIPTION

DATASET SOURCE

The dataset used in this project is obtained from the NASA Exoplanet Archive, a publicly available and scientifically validated repository that contains confirmed exoplanetary data. The dataset is regularly updated and includes a wide range of planetary, orbital, and stellar parameters necessary for habitability analysis.

DATASET SIZE AND STRUCTURE

- The dataset contains thousands of confirmed exoplanet records
- Each record represents one exoplanet
- Data is structured in tabular format (CSV)
- Rows correspond to planets
- Columns represent planetary and stellar attributes
- This structure makes the dataset suitable for machine learning and statistical analysis.

ATTRIBUTE CATEGORIES

1. **PLANETARY ATTRIBUTES**
   These attributes describe the physical nature of exoplanets:
   - Planet Radius
   - Planet Mass
   - Planet Density
   - Equilibrium Temperature

   Smaller, Earth-sized planets with moderate temperatures are more likely to support life. These attributes are critical for identifying rocky planets and assessing surface conditions**.**
2. **ORBITAL ATTRIBUTES**
   These define the planet's orbit around its host star:
   - Orbital Period
   - Semi-major Axis
   - Orbital Eccentricity

   Orbital parameters help determine whether a planet lies within the habitable zone, where liquid water can exist.
3. **HOST STAR ATTRIBUTES**
   These describe the properties of the parent star:
   - Stellar Temperature
   - Stellar Luminosity
   - Stellar Radius
   - Stellar Spectral Type

Host star properties strongly influence planetary climate and radiation exposure. Stars with stable luminosity and moderate temperatures (G and K types) are more favorable for habitability.

DATA QUALITY ANALYSIS

MISSING VALUES

- Several attributes contain missing values due to observational limitations
- Missing data is more common in mass, density, and metallicity

HANDLING STRATEGY

- Remove records with critical missing values
- Use imputation techniques where appropriate

DATA DISTRIBUTION ANALYSIS

- Planet radius distribution is skewed toward larger gas giants
- Temperature values vary widely across exoplanets
- Most exoplanets orbit M-type stars