NASA Exoplanet Archive Dataset Description

1. Introduction

This dataset is sourced from the NASA Exoplanet Archive and contains detailed astronomical data for confirmed exoplanets and their host stars. The dataset has been filtered using the constraints: default_flag = 1 and discovery facility containing 'TESS'. It provides comprehensive planetary, stellar, and observational parameters that are useful for scientific research, statistical analysis, and machine learning model development.

2. Dataset Overview

Total Records (Rows)	39,315
Total Features (Columns)	289
Numerical Features	255
Integer Features	6
Categorical/Text Features	28
Approximate File Size	86.7 MB

3. Planet Identification Attributes

These attributes uniquely identify each planet and its host system. They include planet name, host star name, designation letter, and record identifiers.

4. Planetary Physical Characteristics

These features describe the physical and orbital properties of planets such as orbital period, radius, mass, semi-major axis, eccentricity, equilibrium temperature, and incident stellar radiation. These parameters are essential for habitability analysis and planetary classification.

5. Stellar Properties

These features describe the host star's characteristics including effective temperature, radius, mass, luminosity, metallicity, and surface gravity. Stellar properties strongly influence planetary climate and orbital behavior.

6. Discovery and Detection Information

This section includes discovery method, year of discovery, observing facility, and publication details. It enables analysis of detection trends and telescope contributions.

7. Observational Statistics

Includes counts of photometric, spectroscopic, and radial velocity observations, which indicate the level of observational support for each planetary record.

8. Applications of the Dataset

- Exoplanet classification (e.g., terrestrial vs gas giants)
- Habitability prediction and environmental modeling
- Supervised and unsupervised machine learning tasks
- Statistical astronomy research
- Stellar–planet interaction analysis
- Trend analysis of discovery methods and facilities

9. Data Considerations

The dataset contains missing values across several parameters and may require data cleaning, normalization, and feature selection before use in predictive models. Duplicate planetary systems may exist due to multiple parameter sets; therefore, using the default_flag ensures consistency in analysis.