User Manual – HAWA HP Web Platform

Hawa Hanan Pacha – "Beyond the Sky"
Exoplanet Exploration Project with Artificial Intelligence

Version 1.0 – October 2025

PRESENTATION

This manual is intended to serve as a guide for users of the HAWA HP platform in using the web-based system for exoplanet prediction and analysis using artificial intelligence models.

HAWA HP is a 100% web-based platform that enables researchers, students, and astronomy enthusiasts to analyze data from NASA space missions (Kepler, K2, TESS) to identify potential exoplanets using advanced machine learning techniques.

The platform does not require local installation and can be accessed from any modern web browser with an internet connection.

CHAPTER I. GETTING STARTED WITH HAWA HP

1. Access to the Platform

1.1. System Requirements

To use HAWA HP you need:

- Updated web browser (Chrome, Firefox, Safari, Edge)
- Stable internet connection
- Ability to download CSV files

1.2. Home Page

Upon accessing the platform, you'll find the welcome screen with the message: "Explore the universe of exoplanets with artificial intelligence."

From this page you have two options:

Option 1: Browse as a Guest

- Immediate access without having to create an account
- Functionality limited to informative content
- Redirects to the Guest Page

Option 2: Browse with Account

- Full access to all features
- Ability to train and save your own models
- Redirects to the Login/Registration module

2. Guest Mode

2.1. Guest Mode Features

Guest mode provides informational and educational access that includes:

- **Introduction to the Platform:**Description of the purpose and capabilities of HAWA HP
- **Educational Information:**explanation of what exoplanets are and their scientific importance
- **Prediction Description:**explanation of how artificial intelligence identifies exoplanets
- **Dynamic Statistics:**Visualization of the total number of exoplanets discovered with graphs by year and mission
- **Interactive Visualization:**star maps and dataset examples
- **Public Datasets:**direct links to NASA data sources
- **Prediction Example:**static form displaying sample data and simulated results

2.2. Limitations of Guest Mode

As a guest you may NOT:

- Make real predictions with your own data
- Train custom models
- Save results or history
- Access the full dashboard

To access these features you must create an account.

3. User Authentication

3.1. Create a New Account

To register on the platform:

Step 1: From the home page, select "Browse with Account"

Step 2: In the authentication module, select the "Register" or "Sign Up" option.

Step 3: Complete the following information:

- Valid email address
- Secure password
- Confirm password

Step 4: Accept the terms and conditions

Step 5: Press the "Create Account" button

Step 6: Check your email if required by the system

3.2. Login

If you already have an account:

Step 1: Select "Browse with Account" from the home page

Step 2: In the authentication module, enter:

- Your email
- Your password

Step 3: Press the "Login" button

3.3. Password Recovery

If you forget your password:

Step 1: On the login screen, select "Forgot your password?" or "Password Recovery."

Step 2: Enter the email address associated with your account

Step 3: Press "Send Recovery Link"

Step 4: Check your email and follow the instructions to reset your password.

CHAPTER II. DASHBOARD NAVIGATION

1. Main Interface

Once authenticated, you'll be taken to the HAWA HP main dashboard. This is your main workspace where you can perform all your analysis and prediction operations.

1.1. Dashboard Structure

The Dashboard features a fixed side menu that remains visible at all times. This menu contains three main modules:

1. **Analytics**(Analytics)
2. **Batch Prediction**(Batch Prediction)
3. **Model Training**(Model Training)

1.2. Navigation between Modules

To switch between modules, simply click the desired option in the side menu. The main content of the page will change to display the features of the selected module.

2. Analytics Module

2.1. Description

The Analytics module displays real-time statistics for the best model available in the system. By default, the metrics for the model with the highest accuracy (precision) are displayed.

2.2. Available Metrics

The system displays the following performance metrics:

Accuracy: Proportion of correct predictions out of the total number of predictions made. A value of 0.95 means that the model is correct 95% of the time.

Precision: Of all the cases the model predicted as exoplanets, how many actually are. High accuracy means few false positives.

Recall (Sensitivity): Of all the real exoplanets in the data, how many was the model able to identify? High recall means the model doesn't miss many real exoplanets.

F1-Score: Harmonic mean of precision and recall. This is useful when seeking a balance between the two metrics. An F1 score close to 1.0 indicates a well-balanced model.

2.3. Interpretation of Metrics

To evaluate the quality of a model:

- **Accuracy > 0.90:**excellent model
- **Accuracy 0.80-0.90:**good model
- **Accuracy 0.70-0.80:**acceptable model
- **Accuracy < 0.70:**model requires improvement

Also consider the balance between precision and recall depending on your use case:

- If you prefer to avoid false positives (not misclassify objects as exoplanets): prioritize high precision
- If you prefer not to miss real exoplanets: prioritize high recall

2.4. Exporting Reports

The Analytics module allows you to export the displayed metrics in report format.

To export a report:

Step 1: Make sure you are viewing the metrics for your desired model

Step 2: Locate the export button (usually labeled "Export" or "Download Report")

Step 3: Select the desired format (if options are available)

Step 4: The system will generate and download the file with the metrics

---

CHAPTER III. EXOPLANET PREDICTION

1. Batch Prediction Module

1.1. Description

The Batch Prediction module allows for automatic predictions on large volumes of astronomical data using pre-trained artificial intelligence models.

1.2. Available Models

The platform offers specialized models for three types of NASA mission datasets:

KOI (Kepler Object of Interest) model: Trained with data from the Kepler mission, the first mission dedicated to searching for exoplanets using the transit method.

TOI (TESS Object of Interest) model: Trained with data from the Transiting Exoplanet Survey Satellite (TESS) mission, the all-sky successor to Kepler.

K2 Model: Trained with data from the K2 mission, Kepler mission extension with modified observing strategy.

2. Data Preparation

2.1. Required File Format

The platform only accepts files in CSV (Comma-Separated Values) format.

CSV file requirements:

- Encoding: UTF-8 (recommended)
- Column separator: comma (,)
- First row: column names
- Subsequent rows: numeric data or text as appropriate

2.2. Data Structure by Dataset Type

IMPORTANT: Each mission type (KOI, TOI, K2) requires a specific column format. You must ensure that your CSV file contains the columns corresponding to the type of model you will be using.

23. NASA Reference Datasets

For the exact structure required, please refer to the official NASA datasets:

KOI Dataset (Kepler):
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative

This dataset contains all the objects of interest identified by the Kepler mission. Examine the available columns to understand what data your file should include.

TOI Dataset (TESS):
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=TO I

This dataset contains the objects of interest for the TESS mission. The columnar structure differs from the KOI dataset.

K2 Dataset:
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=k2pandc

This dataset corresponds to the K2 mission with its own data structure.

2.4. Target Columns

NASA datasets include "target" columns that represent the object's actual classification:

- In the KOI dataset: columns likekoi_dispositioneitherkoi_pdisposition
- In the TOI dataset: columns indicating whether the object is a confirmed candidate
- In the K2 dataset: corresponding classification columns

Important Note: When preparing your CSV file for predictions, you can leave the target columns**empty**either**no data**The system will automatically populate these columns with the predictions generated by the model.

If the target columns contain existing data, the system will ignore it and replace it with the new predictions.

2.5. Downloading Sample Data

To facilitate the preparation of your data:

Step 1: Visit one of the NASA dataset links provided above

Step 2: Explore the data table available in the browser

Step 3: Use the download option on the NASA site to get a sample CSV file

Step 4: Open the downloaded file in a spreadsheet (Excel, Google Sheets, etc.)

Step 5: Examine the column names and data types

Step 6: Use this structure as a template to format your own data

3. Make a Prediction

3.1. Step-by-Step Process

To run an exoplanet prediction:

Step 1: Access the Module From the Dashboard side menu, select "Batch Prediction"

Step 2: Select the Dataset Type Before uploading your file, you must specify what type of dataset you will use:

- Select KOI if your data comes from the Kepler mission
- Select TOI if your data comes from the TESS mission
- Select K2 if your data comes from the K2 mission

This selection is crucial because each model was trained with a specific data structure.

Step 3: Select the Model Choose the model you'll use for the prediction. By default, the system will select the model corresponding to the dataset type chosen in the previous step.

Step 4: Review Format Requirements The system will display an example of the required columns for the selected dataset type. Verify that your CSV file contains these columns.

Step 5: Upload the CSV File Click the file upload area or the "Upload CSV" button.

Navigate to the file location on your computer

Select the prepared CSV file

Confirm the selection

Step 6: Automatic Validation The system will automatically validate:

- That the file is in valid CSV format
- That contains the required columns
- That the data is of the expected type

If there are formatting errors, the system will display a message indicating what needs to be corrected.

Step 7: Run Prediction Once the file is validated, press the "Predict" button.

The system will process the data. Depending on the file size, this can take from seconds to several minutes.

Step 8: Download Results When processing is complete, the system will automatically generate a CSV file with the results.

Download the file by clicking the download button that appears.

3.2. Contents of the Results File

The output CSV file will contain the following columns:

Planet Identifier: Depending on the type of dataset:

- For KOI: columns likekepideitherkoi_name
- For TOI: columns liketoi_nameor corresponding identifier
- For K2: K2 mission-specific identifier

Planet Name: If available in the original data, the official name or designation of the object will be included.

Predicted Class (Predicted Label/Class): The resulting classification, which will generally be one of these categories:

- CONFIRMED: high probability of being an exoplanet
- CANDIDATE: A promising object that requires further confirmation
- FALSE POSITIVE: It is probably not an exoplanet.

Predicted Probability:A numerical value between 0 and 1 that indicates the model's confidence in the prediction:

- Values close to 1.0 indicate high confidence
- Values close to 0.5 indicate uncertainty
- Values close to 0.0 indicate high confidence in the opposing class

3.3. Interpretation of Results

To analyze the prediction results:

Objects with High Probability (> 0.9): These are very promising candidates. The model has high confidence in their classification.

Objects with Medium Probability (0.6 - 0.9): Interesting candidates but who could benefit from further analysis or confirmation with other methods.

Objects with Low Probability (< 0.6): Low confidence in the classification. These could be false positives or require additional data.

Recommendations:

- Prioritize for further investigation those objects with probability > 0.85 classified as CONFIRMED
- Objects classified as CANDIDATE with probability > 0.75 also deserve attention
- Objects with probabilities between 0.4 and 0.6 are in the uncertainty zone and require careful analysis

4. Important Considerations

4.1. Input Data Quality

The accuracy of predictions directly depends on the quality of the data provided:

- **Complete Data:**avoid columns with many missing values
- **Correct Data:**Check that the numerical values are within reasonable ranges
- **Consistent Format:**maintain the same format as NASA reference datasets

4.2. System Limitations

Please note that:

- Predictions are probabilistic, not absolute certainties
- Models work best with data similar to that with which they were trained
- Very low quality or incomplete data will produce unreliable predictions
- The system does not replace detailed scientific analysis, but rather complements it.

CHAPTER IV. MODEL TRAINING

1. Model Training Module

1.1. Description

The Model Training module allows users to train their own custom machine learning models using specific data. This advanced feature is useful for researchers who want to:

- Create specialized models for specific data subsets
- Experiment with different hyperparameter settings
- Improve accuracy for particular use cases

2. Preparation for Training

2.1. Required Training Data

To train a model you need:

CSV File with Training Data:

- Must contain the same column structure as official NASA datasets
- **Must include target columns with known values**(This is different from when you make predictions)
- The data must be balanced between the different classes (CONFIRMED, CANDIDATE, FALSE POSITIVE)

2.2. Selecting the Dataset Type

Before starting training, you must specify what type of data you will use:

Step 1: Access the Module From the side menu, select "Model Training"

Step 2: Select Dataset Type Choose between:

- **KOI**for Kepler mission data
- **TOI**for TESS mission data
- **K2**for K2 mission data

This selection determines:

- What columns should your file contain?
- What preprocessing will the system apply?
- Which algorithms will be most appropriate

3. Hyperparameter Configuration

3.1. What are Hyperparameters?

Hyperparameters are settings that control the model's learning process. Unlike the parameters the model learns from data, hyperparameters must be set before training.

3.2. Hyperparameter Configuration in HAWA HP

The system allows you to configure custom hyperparameters to optimize model performance based on your specific needs.

The available hyperparameters vary depending on the machine learning algorithm used, but typically include:

Common examples of hyperparameters:

- Number of trees in ensemble models
- Maximum depth of decision trees
- Learning rate
- Number of neurons in hidden layers (for neural networks)

- Regularization

To configure hyperparameters:

Step 1: In the Model Training module, locate the hyperparameter configuration section

Step 2: Review the default values suggested by the system

Step 3: Modify the values according to your criteria or experience

Step 4: If you are unsure, use the default values that have been optimized for general cases

4. Training Process

4.1. Start Training

Once the parameters are configured:

Step 1: Load Training Data Click "Upload Training Data"

Select your prepared CSV file with the complete target columns

The system will validate that the file contains appropriate data for training.

Step 2: Review Settings Verify that all parameters are configured correctly:

- Selected dataset type (KOI, TOI, or K2)
- Configured hyperparameters
- Data file uploaded

Step 3: Start Training Press the "Train Model" button

The system will begin the training process

Step 4: Monitor Progress During training, the system can display:

- Progress bar
- Partial metrics
- Estimated time remaining

Training time varies depending on:

- Dataset size

- Model complexity
- Hyperparameter configuration

4.2. Completion of Training

When the training is over:

Step 1: Review Metrics The system will display the performance metrics of the trained model:

- Accuracy
- Precision
- Recall
- F1-Score

Step 2: Name and Save the Model Give your trained model a descriptive name:

- Use names that clearly identify the purpose of the model
- Examples: "KOI_High_Precision_2025", "TOI_Experimental_Model_v2"

Press "Save Model"

Step 3: Confirmation The system will confirm that the model has been saved successfully.

Your new model will be available for use in the Batch Prediction module.

5. Evaluation of Trained Models

5.1. Model Comparison

To determine if your custom model improves results:

Method 1: Compare Metrics Compare your model metrics to the system default model in the Analytics module

Method 2: Cross Validation Use different subsets of data for training and testing

Method 3: Testing with Real Data Make predictions with both models on the same dataset and compare results

5.2. Evaluation Criteria

A trained model is successful when:

- Accuracy is equal to or greater than the default model
- Precision and Recall are balanced according to your needs
- F1-Score shows significant improvement

- Predictions on test data are consistent

5.3. Iteration and Improvement

If your model does not meet the expected performance:

Step 1: Review the quality of the training data

- Check that there are enough examples of each class
- Make sure the data is clean and error-free

Step 2: Tune the hyperparameters

- Experiment with different settings
- Document the results of each experiment

Step 3: Consider more training data

- Larger datasets generally produce better models
- Ensure diversity in training examples

Step 4: Train again

- Repeat the process with the improvements implemented
- Systematically compare results

6. Saved Model Management

6.1. Access to Saved Models

Your trained models remain associated with your user account and can be accessed at any time.

To use a saved template:

Step 1: Go to the Batch Prediction module

Step 2: In the model selector, you will find both the system default models and your custom models listed.

Step 3: Select the model you want to use

Step 4: The metrics in the Analytics module will automatically update to show the performance of the selected model.

CHAPTER V. PRACTICAL USE CASES

1. Use Case 1: Rapid Analysis with KOI Data

Aim: Identifying exoplanet candidates from a set of Kepler observations

Scenery: An astronomy student has a CSV file with 500 Kepler observations and wants to quickly identify which ones could be confirmed exoplanets.

Procedure:

Step 1: Data Preparation

- Visit the NASA KOI dataset link
- Download sample data to understand the structure
- Verify that your CSV file contains the same columns
- Make sure the target columns are empty (since you want predictions)

Step 2: Access the Platform

- Log in to HAWA HP with your account
- Navigate to the Batch Prediction module from the side menu

Step 3: Configuration

- Select "KOI" as the dataset type
- The system will automatically select the optimized KOI model
- Review the format example displayed by the system

Step 4: Loading and Prediction

- Upload your CSV file of 500 observations
- Press "Predict"
- Please wait while the system processes (approximately 30 seconds for 500 records)

Step 5: Results Analysis

- Download the CSV file of results
- Open it in a spreadsheet
- Filter by objects with:
    - Predicted Label = "CONFIRMED"
    - Predicted Probability > 0.85
- Identify approximately 15-25 promising candidates

Step 6: Documentation

- Export the metrics report from the Analytics module
- Document the accuracy of the model used (typically > 0.90 for KOI)

Expected Result: Prioritized list of high-confidence exoplanet candidates, ready for further research or presentation in an academic project.

2. Use Case 2: Specialized Training with TOI Data

Aim: Create a custom model focused on hot Jupiter-type exoplanets

Scenery:A researcher wants to improve the detection of giant exoplanets with short orbital periods (hot Jupiters) using TESS data.

Procedure:

Step 1: Specialized Dataset Preparation

- Download the full TOI dataset from NASA
- Filter the data to include only objects with hot Jupiter characteristics:
    - Orbital period < 10 days
    - Planetary radius > 0.8 Jupiter radii
- Keep target columns with known values
- Ensure class balance (confirmed objects vs. false positives)

Step 2: Access the Training Module

- Log in to HAWA HP
- Navigate to "Model Training" from the side menu

Step 3: Initial Setup

- Select "TOI" as the dataset type
- Name the model: "TOI_Hot_Jupiter_2025"

Step 4: Hyperparameter Configuration

- If you have experience with machine learning:
    - Increase model depth to capture complex relationships
    - Adjust regularization to avoid overfitting
- If you are a beginner:
    - Use the default values
    - Document this configuration for future reference

Step 5: Data Loading and Training

- Load your prepared specialized dataset

- Verify that the system confirms the validity of the data
- Start the workout (may take 5-15 minutes)

Step 6: Initial Assessment

- Review the generated metrics:
  - Target: Precision > 0.88 to reduce false positives
  - Acceptable recall: > 0.80
- If the metrics are unsatisfactory, adjust hyperparameters and retrain

Step 7: Validation with Test Data

- Prepare a TOI dataset that you did NOT use for training
- Use your new model in Batch Prediction
- Compare results with the default TOI model

Step 8: Iteration (if necessary)

- If your model does not exceed the default:
  - Check class balance in training data
  - Consider including more examples
  - Tune hyperparameters based on the results

Expected Result: Specialized model that identifies hot Jupiters more accurately than the general model, useful for specific studies of this type of exoplanet.

3. Use Case 3: Comparative Analysis between Missions

Aim: Compare the effectiveness of different models for identifying exoplanets in multiple datasets

Scenery:A research team has data from all three missions (Kepler, TESS, K2) and wants to determine which model provides the best results for their analysis.

Procedure:

Step 1: Dataset Preparation

- Download representative samples of each mission:
  - 1000 KOI records
  - 1000 TOI records
  - 1000 K2 records
- Ensure that each dataset has:
  - Correct format according to your mission
  - Similar proportion of each class
  - Empty target columns for prediction

Step 2: Running KOI Predictions

- Log in to HAWA HP
- Go to Batch Prediction
- Select KOI model
- Load the Kepler dataset
- Download the results file

Step 3: Running TOI Predictions

- Repeat the process by selecting TOI model
- Load the TESS dataset
- Download results

Step 4: Running K2 Predictions

- Repeat the process by selecting the K2 model
- Load the K2 dataset
- Download results

Step 5: Metric Analysis by Model

- Access the Analytics module
- For each model, document:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
- Export metrics reports

Step 6: Prediction Results Analysis For each results file:

- Count how many objects were classified as CONFIRMED
- Calculate the average of predicted probabilities
- Identify how many objects have probability > 0.90

Step 7: Systematic Comparison Create a comparison table:

```None
● Model | Accuracy | Precision | Recall | F1-Score | Confirmed
  (prob>0.90)
● KOI      | [value]      | [value]       | [value]| [value]        | [number]
● TOI      | [value]      | [value]       | [value]| [value]        | [number]
```

```
K2         | [value] | [value]                | [value]| [value] | [number]
```

Step 8: Conclusions

- Identify which model has better overall accuracy
- Determine which model is more conservative (higher precision)
- Identify which model captures more exoplanets (higher recall)
- Decide which model to use based on your research priorities

Expected Result:A deep understanding of the strengths and weaknesses of each model, enabling informed selection of the appropriate model for future analysis based on the type of data and scientific objectives.

CHAPTER VI. PROBLEM SOLVING

1. Common File Upload Problems

1.1. Error: "Invalid File Format"

Problem: The system rejects your CSV file indicating invalid format.

Common Causes:

- The file is not in actual CSV format (may be XLS or XLSX)
- Incorrect separators (semicolons instead of commas)
- Incompatible character encoding
- Corrupt file

Solutions:

Solution 1: Check Actual Format

- Open the file in a plain text editor (Notepad, TextEdit)
- Check that the data is separated by commas
- If you see tabs or semicolons, convert the file

Solution 2: Save As CSV Correct

- Open the file in Excel or Google Sheets
- Select "Save As" or "Download"

- Explicitly choose "CSV (Comma delimited)" or "CSV (comma separator)"
- Save with UTF-8 encoding if the option is available

Solution 3: Verify Integrity

- Try opening the file in different programs
- If the file does not open, it is corrupted
- Regenerate the file from the original source

1.2. Error: "Missing or Incorrect Columns"

Problem: The system indicates that required columns are missing or the names do not match.

Common Causes:

- Column names do not exactly match the expected format
- Essential columns are missing
- Selected dataset type does not match the file

Solutions:

Solution 1: Check Dataset Type

- Make sure you have selected the correct type (KOI, TOI, or K2)
- If you have Kepler data, you must select KOI
- If you have TESS data, you must select TOI
- If you have K2 data, you must select K2

Solution 2: Review Column Names

- Open the corresponding NASA reference dataset
- Compare column names exactly
- Names are case sensitive
- They should not have extra spaces before or after

Solution 3: Complete Missing Columns

- Download a sample file from the NASA site
- Identify which columns are missing in your file
- Add the missing columns to your file
- If you don't have data for a column, leave the cells empty but include the column

1.3. Error: "File Too Large"

Problem: The system rejects the file because it exceeds the maximum allowed size.

Solutions:

Solution 1: Split the Dataset

- Separate your file into multiple smaller files
- Process each file separately
- Combine the results manually afterwards

Solution 2: Reduce Logs

- If you are testing, use a representative sample
- Randomly select a subset of rows
- For complete analysis, consider batch processing

Solution 3: Contact Support

- If you need to process the entire file
- Check for options for users with special needs

2. Prediction Problems

2.1. Problem: "Inconsistent or Strange Predictions"

Symptoms:

- All objects classified in the same category
- Extremely low probabilities for all objects
- Results that contradict known classifications

Possible Causes:

- Very low quality input data
- Many missing values in critical columns
- Incorrect dataset type selected
- Values outside expected ranges

Solutions:

Solution 1: Data Validation

- Review basic statistics of your data:
  - How many empty cells are there?
  - Are the numerical values in reasonable ranges?
  - Are there extreme outliers?

Solution 2: Data Cleanup

- Remove rows with too many missing values (>30% empty columns)
- For occasional missing values, consider:

    - Use average values from the dataset
    - Delete that specific column if it is not critical
- Normalize extreme values

Solution 3: Check Dataset Type

- Confirm that the selected type corresponds to your data
- A KOI file processed with a TOI model will give incorrect results

Solution 4: Use Reference Data

- Download a small official NASA dataset
- Process it with the system
- If it works correctly, the problem is in your data
- Compare the structure of both files

2.2. Problem: "Prediction Process Takes Too Long"

Symptoms:

- Processing takes more than 5 minutes for small datasets
- The page appears frozen
- No indication of progress

Solutions:

Solution 1: Check Connection

- Check your internet connection
- Try refreshing the page if the time exceeds the expected
- Consider a more stable network if this is a recurring problem

Solution 2: Reduce Size

- If the file has more than 10,000 records, split it
- Process in smaller batches

Solution 3: Wait Patiently

- Very large datasets require real-time processing
- Approximate time: 1-2 seconds per 100 records
- Do not close the browser window during the process

Solution 4: Try Again

- If after 10 minutes there is no response
- Refresh the page and try a smaller file

- If the problem persists, there may be a temporary server issue

3. Model Training Problems

3.1. Problem: "Training Fails or Does Not Start"

Causes and Solutions:

Cause 1: Training Data Without Target

- **Verify:**Make sure the target columns have values
- For training, classification columns MUST have data
- Check that at least 90% of the rows have a complete target.

Cause 2: Extreme Class Imbalance

- **Verify:**Count how many examples there are of each kind
- If a class has <5% of the data, the model will not learn well
- **Solution:**Balance the dataset by including more examples from minority classes

Cause 3: Insufficient Data

- A minimum of 200-300 examples are needed for basic training
- Ideal: >1000 examples well distributed among classes

3.2. Problem: "Trained Model Has Very Low Accuracy"

Symptoms:

- Accuracy < 0.60
- Very low F1-Score
- Model does not outperform random predictions

Diagnosis and Solutions:

Step 1: Review Data Quality

- Calculate percentage of missing values
- If >40% of the cells are empty, the dataset is not useful
- Clean and complete data before retraining

Step 2: Check Class Balance

- Give examples for each class
- If one class has 10x more examples than another:
    - Reduce majority class examples (undersampling)
    - Or duplicate minority class examples (oversampling)

Step 3: Review Hyperparameters

- If you used very extreme custom values, there may be overfitting
- Try default values first
- Gradually adjust one parameter at a time

Step 4: Compare to Baseline

- Use the system default model as a reference
- If the default model also has low accuracy in your data:
  - Your data may be very noisy or ambiguous
  - Consider obtaining additional or better quality data

3.3. Problem: "Model Works in Training but Poorly in Actual Prediction"

Diagnosis: Overfitting

The model memorized the training data instead of learning generalizable patterns.

Solutions:

Solution 1: Use More Training Data

- More diverse data helps the model generalize
- Include examples of different periods or conditions

Solution 2: Increase Regularization

- In hyperparameter configuration
- Increase the regularization parameter
- This penalizes excessively complex models

Solution 3: Reduce Model Complexity

- Reduce tree depth
- Reduce the number of layers or neurons

Solution 4: Cross Validation

- When training, separate your data into:
  - 70% training
  - 15% validation
  - 15% test
- The model should never see the test data during training

4. Session and Account Problems

4.1. Problem: "I Can't Log In"

Solutions:

Solution 1: Verify Credentials

- Make sure you write the email correctly
- Check that there are no extra spaces
- Passwords are case sensitive

Solution 2: Reset Password

- Use the "Forgot your password?" option
- Check your inbox and spam folder
- Follow the recovery link

Solution 3: Verify Account

- If it is a new account, please verify that you have confirmed your email.
- Check if you received an activation email

4.2. Problem: "I lost my trained models"

Possible Causes:

- Session expired
- Server synchronization problems

Solutions:

Solution 1: Verify Session

- Log out completely
- Please log in again
- Models should reappear

Solution 2: Check Model Name

- Check that you are searching for the correct name
- Models are listed alphabetically

Solution 3: Contact Support

- If you definitely lost important models
- Provide approximate date of creation of the model
- The technical team can attempt recovery

5. Problems of Interpretation of Results

5.1. Problem: "I don't understand why an object was classified in a certain way"

Reality: Machine learning models are partial "black boxes." It's not always possible to explain every single decision.

Comprehension Strategies:

Strategy 1: Review Probability

- Probabilities close to 0.5 indicate genuine uncertainty
- These objects are in the boundary zone between classes
- Consider them ambiguous cases that require manual analysis

Strategy 2: Compare with Similar Objects

- Look for other objects with similar characteristics
- Were they classified equally?
- Inconsistencies may indicate noise in the data

Strategy 3: Review Input Data

- Check the characteristics of the specific object
- Are there any missing or anomalous values?
- Extreme values can influence unexpected classification

Strategy 4: External Validation

- Use other analysis methods or tools
- Consult scientific literature on the object
- If it is a known object, compare with official classification

5.2. Problem: "Results Contradict Official Rankings"

Considerations:

Point 1: Models Are Not Perfect

- Even with 95% accuracy, there will be 5% errors
- The model may occasionally be wrong

Point 2: Different Criteria

- Official rankings may use stricter criteria
- They can include information that the model does not have

Point 3: Evolutionary Data

- Classifications change with new evidence
- An object classified as a "candidate" can be confirmed after

To do:

Action 1: Document Discrepancies

- Record cases where the model differs from the official classification
- This is valuable information to improve the model

Action 2: Use as a preliminary filter

- Treat predictions as suggestions
- Use additional analysis for final confirmation
- Don't automatically rule out different predictions

Action 3: Consider Retraining

- If there are many systematic discrepancies
- Retraining with more up-to-date data can help

CHAPTER VII. BEST PRACTICES

1. Data Preparation

Practice 1: Preliminary Validation Before uploading data to the platform:

- Open the file in a spreadsheet
- Visually review the first 50 rows
- Verify that the column names exactly match the NASA datasets

- Calculate basic statistics (average, minimum, maximum) to detect anomalies

Practice 2: Documentation Keep a record of:

- Date the data was collected
- Specific source (NASA link)
- Any transformation applied to the data
- Dataset version if applicable

Practice 3: Versioning

- Save multiple versions of your files
- Use descriptive names: "KOI_original_data_2025-10-05.csv"
- Keep backups before modifying data

2. Use of Models

Practice 1: Getting Started with Default Models

- For new users, please use the system models first.
- Familiarize yourself with expected results
- This establishes a baseline for comparison

Practice 2: Systematic Validation When trying a new model:

- Use a known dataset first
- Compare results with official rankings
- Document accuracy in validation data

Practice 3: Conservative Interpretation

- Do not overinterpret results from a single model
- Use multiple pieces of evidence before drawing definitive conclusions
- Consider predicted probabilities, not just classifications

3. Model Training

Practice 1: Quality Data Over Quantity

- 500 well-curated examples are better than 5000 noisy ones
- Invest time in data cleaning
- Manually verify random samples

Practice 2: Documented Experimentation When training models:

- Document each hyperparameter setting
- Record the metrics obtained
- Note observations about the model's behavior

Practice 3: Independent Validation

- Never evaluate a model with data that you used for training
- Always separate a test data set
- Use data from different periods if possible

4. Interpretation of Results

Practice 1: Scientific Context

- Predictions are tools, not absolute truths
- Always consider the astronomical context
- Consult relevant literature for objects of interest

Practice 2: Transparency in Limitations When reporting results:

- Mention the accuracy of the model used
- Indicate the size of the analyzed dataset
- Recognize uncertainties and limitations

Practice 3: Triangulation

- Use multiple methods whenever possible
- Compare with analysis of other instruments
- Consider additional evidence beyond the model predictions

CHAPTER VIII. FREQUENTLY ASKED QUESTIONS

1. General Questions

What is HAWA HP?

HAWA HP (Hawa Hanan Pacha - "Beyond the Sky") is a web platform that uses artificial intelligence to analyze astronomical data and predict the probability that observed celestial objects are exoplanets. The platform processes data from NASA's Kepler, K2, and TESS missions.

Do I need to install any software?

No. HAWA HP is 100% web-based. You just need:

- A modern, updated web browser
- Stable internet connection
- Ability to upload and download CSV files

Is the platform free to use?

Please see the current terms and conditions on the website. The document does not specify the pricing model.

What level of knowledge do I need?

For basic use (predictions with default models):

- Basic knowledge of CSV files
- Fundamental understanding of what exoplanets are

For advanced use (model training):

- Intermediate knowledge of statistics
- Understanding machine learning concepts
- Experience with astronomical data analysis

2. Questions about Data

Where do I get the data to analyze?

Official data are available at:

- Kepler (KOI): https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cumulative
- TESS (TOI): https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=TOI
- K2: https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=k2pandc

Can I use my own observational data?

Only if you format them exactly like the corresponding NASA datasets. Models are trained for specific structures and won't correctly process data with different formats.

What do I do if I have missing values in my data?

For predictions: You can leave empty cells occasionally, but too many missing values (>30%) will reduce reliability.

For training: Data should be as complete as possible, especially target columns should have values for >90% of the rows.

How much data do I need at least?

For prediction: no minimum, can predict from 1 to thousands of objects.

For training: recommended minimum 300-500 examples, ideal >1000 examples well balanced between classes.

3. Questions about Models

How accurate are the models?

Default models typically achieve:

- Accuracy: >0.90 (90%+)
- Precision and Recall: >0.85
- F1-Score: >0.87

Actual accuracy depends on the quality of the input data.

What algorithms are the models built with?

The document does not specify the exact algorithms. Typically, the following are used for this type of problem:

- Random Forests
- Gradient Boosting
- Neural Networks
- Ensemble Methods

Do models improve over time?

The system's default models are periodically updated with new confirmed data from NASA missions. Your custom models remain as you trained them unless you retrain them with new data.

Can I share my trained models with other users?

The document doesn't specify this feature. The templates are associated with your user account.

4. Questions about Results

What does a probability of 0.75 mean?

This means that the model estimates a 75% probability that the object belongs to the predicted class. This is a moderate-high level of confidence, but not definitive.

For reference:

- 0.90: high confidence

- 0.70-0.90: moderate-high confidence
- 0.50-0.70: low confidence
- <0.50: the model favors the opposite class

Why are some known objects classified incorrectly?

Several possible reasons:

1. The model has ~90-95% accuracy, not 100%
2. Input data may be noisy or incomplete
3. The object may be in the boundary zone between classes
4. Official classifications use additional information not available in the dataset

Can I use these results in scientific publications?

The results can be used as a preliminary analysis or initial filter. For scientific publication:

- Mention that you used HAWA HP and specify the model
- Include model metrics (accuracy, etc.)
- Complement with additional analysis and validation
- Follow standard exoplanet confirmation protocols

What is the difference between CONFIRMED, CANDIDATE and FALSE POSITIVE?

- **CONFIRMED**: objects with a high probability of being real exoplanets, typically have passed multiple validations
- **CANDIDATE**: promising objects that require further confirmation before final classification
- **FALSE POSITIVE**: objects that initially appeared to be exoplanets but are probably other phenomena (binary eclipses, instrumental noise, etc.)

5. Technical Questions

Are there limits on file size?

The document doesn't specify exact limits. If you encounter restrictions, split your dataset into smaller files and process them in batches.

How long does a prediction take?

Approximately:

- 100 records: 1-2 seconds
- 1000 records: 10-20 seconds
- 10000 records: 1-3 minutes

Time varies depending on server load and model complexity.

How long does it take to train a model?

It depends on the size of the dataset and its complexity:

- Small dataset (500 records): 2-5 minutes
- Medium dataset (2000 records): 5-15 minutes
- Large dataset (10000+ records): 15-45 minutes

Is my data saved?

Trained models are associated with your account. Regarding the datasets you upload for prediction, please refer to the platform's privacy policy.

Can I use the platform from mobile devices?

HAWA HP is a web application that runs in browsers. It can be accessed from mobile devices, but the experience is optimized for desktop computers due to the complexity of CSV file handling and operations.

6. Common Problems

I don't see my saved models after training

Solution: Log out completely and log back in. The models should appear in the Batch Prediction module selector.

The metrics displayed do not change when I select a different model

Solution: The Analytics module displays the most accurate model by default. In Batch Prediction, selecting a specific model should update the metrics. If they don't, refresh the page.

My file was rejected but I know it's correct.

Verify:

1. Did you select the correct dataset type (KOI, TOI, K2)?
2. Do the column names match EXACTLY the reference dataset?
3. Is the file an actual CSV and not Excel saved as CSV?

The results seem random

Probable causes:

- Selected dataset type does not match the data
- Very low quality data with many missing values
- You selected the wrong model for the data type

I can't download the results

Solution: Check your browser settings to allow downloads. Check your default download folder. Try a different browser if the problem persists.

CONTACT AND SUPPORT INFORMATION

For additional help, to report issues, or to make suggestions, please visit the HAWA HP contact page on the platform website.

Additional Resources:

- Official NASA documentation on exoplanets:https://exoplanets.nasa.gov/
- NASA Exoplanet Archive:https://exoplanetarchive.ipac.caltech.edu/
- Tutorials on machine learning for astronomy
- Scientific community searching for exoplanets

Last Update: October 2025 **Manual Version:**1.0

HAWA HP – Hawa Hanan Pacha
Exploring the universe beyond the sky

-

Informative Content Guide for Guest Mode

Version 1.0 – October 2025

PRESENTATION

This manual documents the educational and informative content that the HAWA HP chatbot provides to users in guest mode. The purpose is to offer clear, accessible, and scientifically accurate information about exoplanets and the use of artificial intelligence for their detection.

CHAPTER I. INTRODUCTION TO THE PLATFORM

1. What is HAWA HP?

HAWA HP (Hawa Hanan Pacha) is a Quechua expression meaning "Beyond the Sky." The platform is a web-based system that uses artificial intelligence to analyze astronomical data from NASA space missions and predict the probability that observed celestial objects are exoplanets.

2. Purpose of the Platform

HAWA HP has three main objectives:

Objective 1: Democratize Astronomical Analysis To make advanced data analysis tools available to researchers, students, and astronomy enthusiasts without the need for complex programming or expensive computing resources.

Goal 2: Accelerate Discovery Automate the analysis of large volumes of astronomical data that would take months or years to review manually, reducing the time to identify exoplanet candidates from weeks to minutes.

Objective 3: Education and Outreach Provide an educational platform that explains complex scientific concepts in an accessible way and enables hands-on experimentation with real-world data from space missions.

3. Main Capabilities

Automated Prediction: Analysis of astronomical data using machine learning models trained on thousands of confirmed observations from the Kepler, K2, and TESS missions.

Specialized Models: Three types of models optimized for different space missions:

- KOI model for Kepler data
- TOI model for TESS data
- K2 model for Kepler's extended mission

Personalized Training: Ability for advanced users to train their own models with specific settings and custom datasets.

Metrics Analysis: Real-time visualization of model performance using standard machine learning metrics (accuracy, precision, recall, F1-score).

Batch processing: Ability to analyze from a single object to thousands of observations simultaneously.

4. Target Audience

Astronomical Researchers: Scientists who need to sift through large volumes of data to identify promising candidates that merit detailed analysis.

University Students: Students in astronomy, physics, data science, or related fields who want hands-on experience with real-world data and applied machine learning techniques.

Educators: Teachers looking for tools to teach modern astronomy concepts, data analysis, and artificial intelligence applications.

Amateur Astronomers: Enthusiasts with basic knowledge who want to contribute to the field of exoplanet research or learn about modern detection methods.

CHAPTER II. EDUCATIONAL INFORMATION ABOUT EXOPLANETS

1. What are Exoplanets?

Definition: An exoplanet (or extrasolar planet) is any planet orbiting a star other than the Sun. These worlds exist outside our Solar System and represent one of the most active fields of contemporary astronomical research.

Historical Context: Although its existence was theorized for centuries, the first confirmed exoplanet orbiting a Sun-like star was discovered in 1995 by astronomers Michel Mayor and Didier Queloz (2019 Nobel Prize in Physics). Since then, more than 5,000 exoplanets have been confirmed.

2. Scientific Importance

Understanding Planetary Formation: Exoplanets provide crucial insights into how planetary systems form and evolve. The observed diversity (giant planets close to their stars, multi-planet systems, eccentric orbits) has both challenged and refined our theories of planetary formation.

Search for Extraterrestrial Life: Some exoplanets orbit in their stars' "habitable zones," where temperatures could allow liquid water to survive on the surface. Studying these worlds is essential to answering one of humanity's deepest questions: Are we alone in the universe?

Solar System Context: Comparing our Solar System with other planetary systems helps us understand how common or unique our planetary configuration is and what factors may have favored the development of life on Earth.

Atmospheric Physics and Composition: The study of exoplanetary atmospheres allows us to investigate physical and chemical processes in conditions very different from those on Earth, expanding our knowledge of planetary physics.

3. Types of Exoplanets

Exoplanets are generally classified according to their size and composition:

Hot Jupiters: Giant gas planets similar to Jupiter but orbiting very close to their stars, with orbital periods of only days. Surface temperatures can exceed 1,000°C. They were the first type of exoplanets detected due to their large size and close orbits.

Neptunes: Planets similar in size to Neptune, typically with rocky cores surrounded by thick atmospheres of hydrogen and helium. They are very common in the galaxy.

Super-Earths: Rocky planets with a mass greater than Earth but less than Neptune (typically 2-10 Earth masses). There is no analogue in our Solar System, but they appear to be abundant in the galaxy.

Terrestrial: Rocky planets of similar size and composition to Earth, Mars, or Venus. They are particularly interesting when found in habitable zones.

Brown Dwarfs (sub-stellar objects): Objects with masses between planets and stars (13-80 Jupiter masses) that are not massive enough for hydrogen nuclear fusion. They are sometimes detected in exoplanet surveys.

4. Detection Methods

Transit Method: It detects the small decrease in a star's brightness when a planet passes in front of it from our perspective. This is the method used by the Kepler, K2, and TESS missions, and the one analyzed by HAWA HP.

Characteristics:

- Typical brightness decrease: 0.01% - 3%
- Requires the planet to cross in front of its star from our perspective
- Allows you to determine the size of the planet
- Repeated observations confirm orbital periodicity

Radial Velocity: It detects the gravitational wobble that a planet induces on its star. The star moves slightly in response to the planet's gravity, causing Doppler shifts in its spectrum.

Direct Image: Direct photography of the planet separated from its star. Extremely difficult due to the star's brightness, but possible for young, hot planets far from relatively nearby stars.

Gravitational Microlensing: It detects the gravitational lensing effect that a star-planet system exerts on the light from a more distant background star. It allows for the detection of very distant planets.

Astrometry: It measures the precise motion of a star in the sky caused by the planet's gravity. It requires extremely precise position measurements.

5. Space Missions

Kepler (2009-2013):

- First mission specifically dedicated to the search for exoplanets
- He continuously observed 150,000 stars in a fixed region of the sky
- Discovered more than 2,600 confirmed exoplanets
- He showed that planets are extremely common in the galaxy
- Dataset: KOI (Kepler Objects of Interest)

K2 (2014-2018):

- Kepler mission extended after mechanical failures
- Observed multiple regions of the sky over periods of ~80 days
- Discovered more than 500 additional exoplanets
- It allowed studies of various types of stars

TESS (2018-present):

- Transiting Exoplanet Survey Satellite, successor to Kepler
- Observe almost the entire sky, prioritizing bright and nearby stars
- Designed to find candidates for follow-up with larger telescopes
- Focus on small planets around nearby stars
- Dataset: TOI (TESS Objects of Interest)

6. Discovery Statistics

Current Issues (October 2025):

- More than 5,500 confirmed exoplanets
- More than 10,000 additional candidates are in the confirmation process
- Discovered planetary systems: more than 4,000
- Planets in the habitable zone: more than 60 confirmed

Trend in Discoveries by Year:

- 1995-2009: ~350 planets (pre-Kepler era)
- 2009-2013: explosion of discoveries with Kepler
- 2014-2018: Constant discoveries with K2 and other methods
- 2018–present: TESS maintains accelerated pace of discoveries

Distribution by Method:

- Transit: ~75% of discoveries confirmed
- Radial velocity: ~20%
- Other methods: ~5%

Size Distribution:

- Neptunes: ~35%
- Super-Earths: ~30%
- Jupiters: ~25%
- Terrestrial: ~10%

CHAPTER III. ARTIFICIAL INTELLIGENCE AND PREDICTION

1. What is a Prediction?

Definition in Astronomical Context:A prediction, in the context of HAWA HP, is the result of applying artificial intelligence algorithms to observational data to estimate the probability that a detected object is a genuine exoplanet.

Prediction Process:

1. Stellar brightness data is captured for weeks or months
2. Patterns of periodic decrease in brightness are detected
3. Numerical characteristics are extracted from these observations
4. The AI model evaluates these characteristics
5. A classification and associated probability are generated

Difference between Detection and Prediction:

- **Detection**: Automatic or manual identification of traffic signs from brightness data
- **Prediction**: probabilistic assessment of whether that signal corresponds to a real exoplanet

2. Why use Artificial Intelligence?

Data Volume Problem: Modern space missions generate massive amounts of data:

- Kepler observed 150,000 stars simultaneously
- TESS observes 200,000 stars per sector
- Each star generates thousands of brightness measurements
- Manually reviewing all this data would take decades

False Positives Challenge: Not all decreases in brightness indicate exoplanets. Alternative causes include:

- Eclipsing binary stars
- Natural stellar variability
- Rotating starspots
- Instrumental noise
- Light pollution from neighboring stars

AI models learn to distinguish real exoplanets from these phenomena.

Advantages of AI:

- **Speed**: Analyze thousands of objects in minutes vs. months of manual analysis
- **Consistency**: applies the same criteria uniformly
- **Scalability**: can process increasingly larger datasets without a proportional increase in resources

- **Subtle Pattern Detection**: identifies complex correlations that humans might overlook

3. How does the process work?

Step 1: Data Collection Space telescopes measure the brightness of stars every 2 minutes (TESS) or 30 minutes (Kepler), generating light curves that show how the brightness changes over time.

Step 2: Preprocessing The raw data is cleaned to remove:

- Instrumental effects
- Variations caused by the rotation of the satellite
- Systematic trends unrelated to planetary transits

Step 3: Event Detection Automatic algorithms identify periodic decreases in brightness that could be planetary transits.

Step 4: Feature Extraction For each detected event, metrics such as:

- Transit depth (how much the brightness decreases)
- Duration of transit
- Orbital period
- Shape of the traffic curve
- Signal-to-noise ratio
- Characteristics of the host star

Step 5: Classification with AI The machine learning model, trained with thousands of known examples, evaluates these characteristics and produces:

- A classification (CONFIRMED, CANDIDATE, FALSE POSITIVE)
- An associated probability (0.0 to 1.0)

Step 6: Human Validation For scientific publication, high-probability candidates are reviewed by astronomers who:

- Examine light curves visually
- They look for additional information in other catalogs
- They carry out follow-up observations
- Additional statistical tests are applied

4. Machine Learning in Astronomy

What is Machine Learning?Machine learning is a subfield of artificial intelligence where algorithms learn patterns from data without being explicitly programmed with fixed rules.

Supervised Learning:HAWA HP uses supervised learning, where:

1. Examples are provided with known answers (confirmed exoplanets vs false positives)
2. The algorithm learns what features distinguish each class
3. The resulting model can classify new examples with no known answer

Common Model Types:

- **Random Forests**: ensembles of decision trees that vote on classification
- **Gradient Boosting**: sequences of models where each one corrects errors of the previous one
- **Neural Networks**: Brain-inspired architectures that learn hierarchical representations

Evaluation Metrics:

Accuracy: Proportion of correct predictions out of the total. Example: an accuracy of 0.92 means 92% correct.

Precision: Of all objects classified as exoplanets, what fraction actually are? High precision minimizes false positives.

Recall (Sensitivity): Of all real exoplanets, what fraction did the model identify? High recall minimizes false negatives (undetected exoplanets).

F1-Score: Harmonic mean of precision and recall. Useful when seeking a balance between the two metrics.

Trade-offs:

- Very conservative models: high precision, low recall (they lose real exoplanets)
- Very permissive models: high recall, low precision (many false positives)
- Balanced models: moderately high precision and recall

5. Limitations of the Models

Limitation 1: Input Data Quality Models can't compensate for very low-quality data. Garbage in, garbage out.

Limitation 2: Generalization Models work best with data similar to what they were trained on. Very unusual objects may be misclassified.

Limitation 3: They are not Oracles Predictions are probabilistic, not certain. An object with a probability of 0.99 still has a 1% chance of being a false positive.

Limitation 4: Partial Black Boxes It is not always possible to explain exactly why a model made a certain decision for a specific object.

Limitation 5: Biases in Training Data If the training data does not adequately represent the diversity of real-life cases, the model will inherit these biases.

6. Why Not Just Use AI?

AI models are powerful but complementary tools to human analysis:

Reason 1: Scientific Confirmation The astronomical community requires multiple lines of independent evidence before definitively confirming an exoplanet.

Reason 2: Extreme Cases Very unusual objects or those at the limits of detection require case-by-case evaluation by experts.

Reason 3: Scientific Context Humans can consider the broader context: Is this system physically plausible? What do we know about the host star?

Reason 4: New Discoveries Completely new phenomena won't be present in training data. Humans can recognize something genuinely novel.

Best Approach: Human-AI Collaboration

- AI filters millions of objects to thousands of promising candidates
- Humans analyze these prioritized candidates in detail
- AI accelerates; humans validate and discover

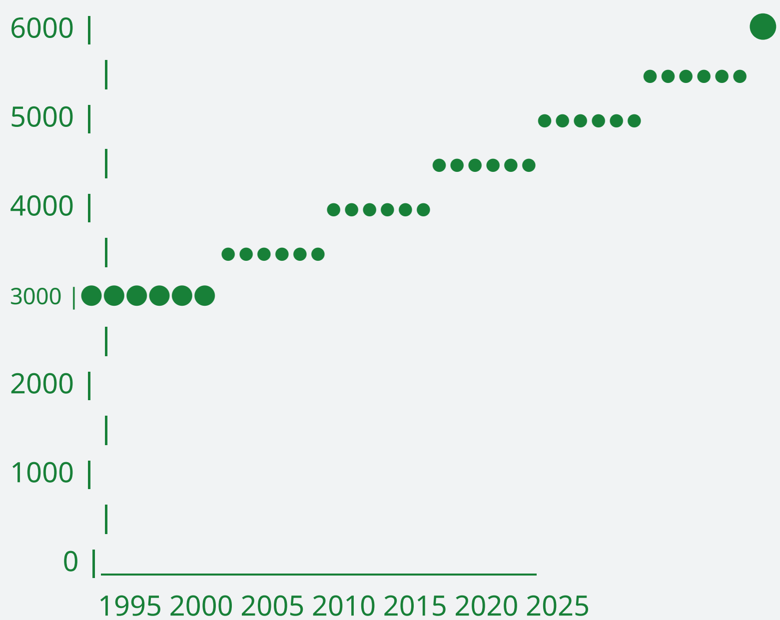
CHAPTER IV. VISUALIZATIONS AND EXAMPLES

1. Displaying Dynamic Statistics

Discovered Exoplanets: Temporal Evolution

Conceptual timeline chart showing cumulative growth of discoveries:

```
None
Confirmed Exoplanets by Year
```

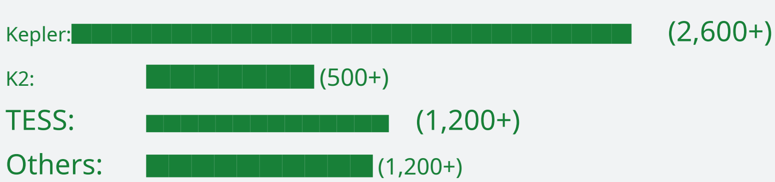


Key events:
1995: First exoplanet (51 Pegasi b) 2009: Kepler launched
2018: TESS mission begins

Distribution by Mission:

None

Discoveries by Space Mission



Types of Exoplanets:

None

Distribution by Category

```
Neptunes:      ███████████████  35%
Super-Earths:                 30%
[illegible] ██████ 25%
Terrestrial:   ████10%
```

2. Star Maps and Interactive Visualization

Concepts for Visualization:

Kepler Sky Map:A specific region in the constellations Cygnus and Lyra that Kepler observed continuously for four years. Approximately 115 square degrees of the sky.

TESS Coverage:TESS observes almost the entire sky, divided into 26 sectors, each observed for approximately 27 days. It covers 85% of the entire sky.

Galactic Distribution: The exoplanets detected are relatively nearby (most within 3,000 light-years). The Milky Way is 100,000 light-years across, so we only explored our local neighborhood.

Interactive Display Elements:

- Points representing analyzed stars
- Colors indicating whether they have confirmed exoplanets, candidates, or none
- Dot size proportional to the number of planets in the system
- Filters by planet type, size, orbital period
- Hover information: system name, distance, planet characteristics

3. Example Datasets

KOI Dataset Structure (Kepler):

Example of typical columns:

```None
kepid: Kepler star identifier koi_name: Name of the
object of interest
```

```
koi_disposition: Classification (CONFIRMED, FALSE POSITIVE, CANDIDATE)

koi_period: Orbital period in days
koi_depth: Transit depth in parts per million koi_duration: Transit duration
in hours
koi_prad: Radius of the planet in Earth radii koi_teq: Equilibrium
temperature of the planet in Kelvin koi_steff: Effective temperature of the
star in Kelvin
```

Example of data row:

```None
kepid: 10593626
koi_name: K00752.01
koi_disposition: CONFIRMED
koi_period: 3.5225
koi_depth: 1520.0
koi_duration: 2.94
koi_prad: 1.52
koi_teq: 1340
koi_steff: 6117
```

TOI Dataset Structure (TESS):

Example of typical columns:

```None
TIC ID: TESS Input Catalog identifier TOI: TESS
Object of Interest number Disposition: Planet
status
Period (days): Orbital period Duration (hours):
Transit duration Depth (ppm): Transit depth

Planet Radius (R_Earth): Planetary Radius Stellar Radius
(R_Sun): Stellar Radius
```

K2 Dataset Structure:

Similar to KOI but with K2-specific identifiers and different fields depending on the observation campaign.

4. Static Prediction Example

Demonstration Scenario:

```
None
SAMPLE ENTRY

Star: K2-18 (K2)
Stellar Temperature: 3503 K Stellar
Radius: 0.41 R_Sol

Traffic Signal Detected:
- Period: 32.94 days
- Depth: 800 ppm
- Duration: 3.2 hours
- Number of transits observed: 8

Derived Features:
- Estimated planetary radius: 2.24 R_Earth
- Equilibrium temperature: 265 K
- Living area: YES
- Signal-to-noise ratio: 15.3

PROCESSING WITH K2 MODEL...
```

```
None
PREDICTION RESULT

Classification: CONFIRMED Probability: 0.94
(94% confidence)

Interpretation:
```

The model indicates a high probability that this object is a genuine exoplanet.
The characteristics are consistent with a super-Earth in the habitable zone of a red dwarf.

Recommendation:
Excellent candidate for follow-up observations, especially atmospheric studies with telescopes such as the James Webb telescope.

Note: This is the real planet K2-18b, confirmed and extensively studied. Water vapor was detected in its atmosphere in 2019.

CHAPTER V. RESOURCES AND LINKS

1. NASA Public Datasets

Kepler Objects of Interest (KOI):URL: https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=cu mulative

Content:

- All Kepler mission objects of interest
- Includes confirmed, candidates, and false positives
- Planetary and stellar characterization data
- Light curves available for download

TESS Objects of Interest (TOI):URL: https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=TO I

Content:

- TESS Mission Objects of Interest
- Regularly updated with new discoveries
- Includes community and TESS team provisions
- Links to tracking data

K2 Targets and Candidates:URL: [https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=k2p andc](https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=k2pandc)

Content:

- K2 Candidates and Confirmed Planets
- Data from multiple observation campaigns
- Planetary and stellar characteristics

2. How to Use Reference Datasets

For Guest Mode Users:

Step 1: Basic Exploration

- Visit any of the links above
- Browse the NASA Exoplanet Archive web interface
- Explore the available columns by clicking on the headings
- Read the descriptions of each field

Step 2: Data Visualization

- Use site filters to view specific subsets
- Example: Filter by koi_disposition = "CONFIRMED" to see only confirmed planets
- Sort by different columns to find interesting records

Step 3: Understanding Magnitudes

- Note the typical ranges of values (periods from hours to years, radii from 0.5 to 20+ R_Earth)
- Identify which fields are most relevant for different types of analysis

For Future Registered Users:

Step 4: Data Download

- Use the "Download Table" button to get the full CSV
- Select specific columns if you only need a subset
- Choose CSV (comma-separated values) format

Step 5: Preparation for HAWA HP

- Open the CSV in a spreadsheet
- Check column names (do not modify)
- This structure is required for loading in HAWA HP

- You can add your own data following this format exactly

3. Additional Educational Resources

NASA Exoplanet Exploration:https://exoplanets.nasa.gov/

NASA's official exoplanet site with:

- News of recent discoveries
- Accessible explanations of concepts
- Videos and animations
- Resources for educators

NASA Exoplanet Archive:https://exoplanetarchive.ipac.caltech.edu/

Central portal for:

- Access to all datasets
- Online analysis tools
- Detailed technical documentation
- Data usage tutorials

Fundamental Concepts:

- Transit method: Search for "transit method animation NASA"
- Habitable zones: search for "habitable zone calculator"
- Exoplanet types: search for "exoplanet types comparison"

4. Glossary of Terms

Astronomical Terms:

Transit: The passage of a planet in front of its star from our perspective, causing a decrease in star brightness.

Light Curve: Graph of a star's brightness as a function of time.

Orbital Period: Time it takes for a planet to complete one orbit around its star.

Planetarium Radio: Planet size, usually expressed in Earth radii (R⊕)or Jupiterians (RJ).

Living Area: Region around a star where a planet could maintain liquid water on its surface.

Equilibrium Temperature: Temperature that a planet would have assuming a balance between received and emitted radiation.

Transit Depth: Fraction of starlight blocked by the planet, typically in parts per million (ppm).

Host Star: The star around which an exoplanet orbits.

Machine Learning Terms:

Model: Mathematical algorithm trained to make predictions based on patterns learned from historical data.

Training: Process of fitting a model using data with known responses.

Prediction: Model exit when it processes new data with no known response.

Hyperparameters: Settings that control the learning process, established before training.

Overfitting: When a model memorizes training data instead of learning generalizable patterns.

Accuracy: Proportion of correct predictions out of total predictions.

Precision: Proportion of positive predictions that are correct.

Recall: Proportion of actual positive cases that were identified.

F1-Score: Harmonic mean of precision and recall.

Classification Terms:

CONFIRMED: Exoplanet validated by multiple independent lines of evidence.

CANDIDATE: Object with promising traffic signal but not yet definitively confirmed.

FALSE POSITIVE:A sign that initially seemed like a planetary transit but turned out to be another phenomenon.

Predicted Probability: Numerical value (0-1) indicating the model's confidence in its classification.

Data Dictionary – NASA Exoplanet Datasets

Column Documentation for KOI, TOI and K2

Version 1.0 – October 2025

PRESENTATION

This document provides detailed definitions of each column present in NASA's public exoplanet datasets. It is intended as a technical reference for HAWA HP users who need to understand the structure and meaning of the data they upload to the platform.

CHAPTER I. DATASET KOI (KEPLER OBJECTS OF INTEREST)

1. Identifiers

kepid

- **Guy:**Whole
- **Description:**Kepler Input Catalog ID. Unique identifier of the star in the Kepler input catalog.
- **Typical range:**757076 - 12644769
- **Example:**10593626
- **Grades:**A star can have multiple KOIs if multiple transit signals were detected.

kepoi_name

- **Guy:**Text
- **Description:**Kepler Object of Interest Name. Unique identifier of the candidate object.
- **Format:**K + 5 digits + period + 2 digits (K00752.01)
- **Example:**K00752.01

- **Grades:**The number after the dot indicates multiple planets in the same system (01, 02, 03...).

kepler_name

- **Guy:**Text
- **Description:**Official name of the exoplanet confirmed.
- **Format:**Kepler-### + letter (Kepler-22b)
- **Example:**Kepler-22b
- **Grades:**Only present for objects with a CONFIRMED disposition. The first confirmed planet in a system receives the letter 'b', the second 'c', etc.

2. Classification and Disposition

koi_disposition

- **Guy:**Categorical text
- **Description:**Official classification of the object.
- **Possible values:**
    - CONFIRMED: Exoplanet validated
    - FALSE POSITIVE: It is not a planet
    - CANDIDATE: Candidate not yet confirmed
- **Example:**CONFIRMED
- **Grades:**This is the final ranking after detailed analysis.

koi_pdisposition

- **Guy:**Categorical text
- **Description:**Preliminary layout based on automated analysis of the Kepler pipeline.
- **Possible values:**
    - CANDIDATE: Passes all automatic tests
    - FALSE POSITIVE: Fails one or more tests
- **Example:**CANDIDATE
- **Grades:**May differ from koi_disposition after manual validation.

koi_score

- **Guy:**Decimal (0.0 - 1.0)
- **Description:**Confidence score from the automated pipeline. Probability that the object is a planet based on automated analysis.
- **Range:**0.0 (definitely not a planet) to 1.0 (most likely a planet)
- **Example:**0.943
- **Grades:**Scores > 0.9 generally indicate a high probability of a real planet.

3. False Positive Flags

koi_fpflag_nt

- **Guy:**Integer (0 or 1)
- **Description:**Not Transit-Like flag. Indicates whether the signal does not resemble a planetary transit.
- **Values:**
  - 0: The signal looks like a transit
  - 1: The sign does not look like a transit
- **Example:**0
- **Grades:**Activated when the light curve shape is inconsistent with a transit.

koi_fpflag_ss

- **Guy:**Integer (0 or 1)
- **Description:**Stellar Eclipse flag. Indicates whether the signal is likely a binary stellar eclipse.
- **Values:**
  - 0: It doesn't look like a binary eclipse
  - 1: Probably binary eclipse
- **Example:**0
- **Grades:**Binary eclipses show secondary transits or characteristic shape variations.

koi_fpflag_co

- **Guy:**Integer (0 or 1)
- **Description:**Centroid Offset flag. Indicates whether the light centroid shifts significantly during transit.
- **Values:**
  - 0: No significant displacement
  - 1: Displacement detected
- **Example:**0
- **Grades:**Shift indicates that the signal is coming from a contaminating background star, not the target star.

koi_fpflag_ec

- **Guy:**Integer (0 or 1)
- **Description:**Ephemeris Match Indicates Contamination flag. The signal matches the orbital period of another known KOI.
- **Values:**
  - 0: No match with another KOI
  - 1: Coincides with the period of another KOI
- **Example:**0

- **Grades:**Suggests contamination of light from neighboring star with real planet.

4. Orbital Parameters

koi_period

- **Guy:**Decimal
- **Units:**Days
- **Description:**Orbital period of the planet. Time between consecutive transits.
- **Typical range:**0.5 days (ultra-hot Jupiters) to 700+ days (planets in extended orbits)
- **Example:**289,862
- **Grades:**Very short periods (<1 day) or very long periods (>500 days) are less reliable.

koi_period_err1

- **Guy:**Decimal
- **Units:**Days
- **Description:**Positive error (top) in the measurement of the orbital period.
- **Example:**+0.025
- **Grades:**Asymmetric uncertainty is common in astronomy.

koi_period_err2

- **Guy:**Decimal
- **Units:**Days
- **Description:**Negative (lower) error in the measurement of the orbital period.
- **Example:**-0.025
- **Grades:**Negative value indicates downward error.

koi_time0bk

- **Guy:**Decimal
- **Units:**BKJD (Barycentric Kepler Julian Date)
- **Description:**Time of the first observed transit. Reference period for calculating future transits.
- **Typical range:**100 - 1600 (days since the start of the Kepler mission)
- **Example:**170.5387
- **Grades:**BKJD is a Kepler-specific timescale.

koi_time0bk_err1

- **Guy:**Decimal
- **Units:**Days
- **Description:**Positive error in the time of transit.
- **Example:**+0.0012

koi_time0bk_err2

- **Guy:**Decimal
- **Units:**Days
- **Description:**Negative error in the transit period.
- **Example:**-0.0012

5. Transit Geometry

koi_impact

- **Guy:**Decimal
- **Description:**Impact parameter. Minimum distance from the center of the planet to the center of the star during the transit, in units of stellar radii.
- **Range:**0.0 (central traffic) to ~1.0 (low traffic)
- **Example:**0.146
- **Grades:**Values close to 0 = central transits (maximum duration). Values close to 1 = low-level transits (minimum duration, difficult to detect).

koi_impact_err1

- **Guy:**Decimal
- **Description:**Positive error of the impact parameter.
- **Example:**+0.034

koi_impact_err2

- **Guy:**Decimal
- **Description:**Negative error of the impact parameter.
- **Example:**-0.034

koi_duration

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Transit duration. Total time from start to finish of the transit.
- **Typical range:**0.5 - 12 hours
- **Example:**2.940
- **Grades:**Duration depends on the size of the star, the orbital speed of the planet, and the impact parameter.

koi_duration_err1

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Positive error of transit duration.

- **Example:**+0.056

koi_duration_err2

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Negative error of transit duration.
- **Example:**-0.056

6. Transit Characteristics

koi_depth

- **Guy:**Decimal
- **Units:**ppm (parts per million)
- **Description:**Transit depth. Fraction of starlight blocked by the planet.
- **Typical range:**10 ppm (small planets) to 30,000 ppm (Jupiters)
- **Example:**1520.0
- **Grades:**Depth$\propto$ (Rplanet/Rstar)^2. Detecting transits requires depth > instrumental noise (~20 ppm for Kepler).

koi_depth_err1

- **Guy:**Decimal
- **Units:**ppm
- **Description:**Positive error of transit depth.
- **Example:**+45.3

koi_depth_err2

- **Guy:**Decimal
- **Units:**ppm
- **Description:**Negative error of transit depth.
- **Example:**-45.3

7. Derived Planetary Properties

koi_prad

- **Guy:**Decimal
- **Units:**Terrestrial radios (R⊕)
- **Description:**Planetary radius. Calculated from the transit depth and stellar radius.
- **Typical range:**0.5 R⊕ (sub-Earths) to 20+ R⊕ (Jupiters)
- **Example:**1.52
- **Grades:**1 R⊕ =Earth's radius. ~11.2 R⊕ =Jupiter's radius.

koi_prad_err1

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Positive error of the planetary radius.
- **Example:**+0.18

koi_prad_err2

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Negative error of the planetary radius.
- **Example:**-0.18

koi_teq

- **Guy:**Decimal
- **Units:**Kelvin (K)
- **Description:**Equilibrium temperature of the planet. Temperature assuming balance between radiation received and emitted.
- **Typical range:**200 K (cool planets) to 3000 K (ultra-hot Jupiters)
- **Example:**1340
- **Grades:**Assumes albedo = 0 and perfect heat redistribution. Actual temperature may differ.

koi_teq_err1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive error of the equilibrium temperature.
- **Example:**+48

koi_teq_err2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative error of the equilibrium temperature.
- **Example:**-48

koi_insol

- **Guy:**Decimal
- **Units:**Overland flow (F⊕)
- **Description:**Insolation. The amount of radiation received by the planet relative to that received by the Earth from the Sun.
- **Typical range:**0.1 F⊕ (cold planets) at 10,000+ F⊕ (ultra-hot)

- **Example:**56.3
- **Grades:**1.0 = overland flow. Conservative habitable zone: 0.35-1.7 F⊕.

koi_insol_err1

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Positive insolation error.
- **Example:**+3.2

koi_insol_err2

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Negative insolation error.
- **Example:**-3.2

8. Detection Metrics

koi_model_snr

- **Guy:**Decimal
- **Description:**Signal-to-Noise Ratio of the traffic model. The ratio of signal depth to background noise.
- **Typical range:**7 (detection limit) to 100+ (very clear detections)
- **Example:**34.2
- **Grades:**SNR > 7.1 is a typical threshold for reliable detection. SNR > 20 indicates robust detection.

koi_tce_plnt_num

- **Guy:**Whole
- **Description:**Planet number in the Threshold Crossing Event (TCE). Order in which the signals were detected in the system.
- **Example:**1
- **Grades:**TCE is the initial event when the signal crosses the auto-detection threshold.

koi_tce_delivname

- **Guy:**Text
- **Description:**Name of the TCE catalog delivery where the object was reported.
- **Example:**q1_q17_dr25_tce
- **Grades:**Indicates which quarters of Kepler data were used in the analysis.

9. Stellar Properties

koi_steff

- **Guy:**Decimal
- **Units:**Kelvin (K)
- **Description:**Effective temperature of the host star.
- **Typical range:**3000 K (M dwarfs) to 10,000 K (A-type stars)
- **Example:**6117
- **Grades:**Sun = 5778 K. Temperature determines the color and spectral type of the star.

koi_steff_err1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive error of stellar temperature.
- **Example:**+82

koi_steff_err2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative error of the stellar temperature.
- **Example:**-82

koi_slogg

- **Guy:**Decimal
- **Units:**$\log_{10}(cm/s^2)$
- **Description:**Logarithm of the surface gravity of the star.
- **Typical range:**3.5 (giants) to 5.0 (dwarfs)
- **Example:**4.35
- **Grades:**Sun = 4.44. Smaller log g = larger/more evolved star.

koi_slogg_err1

- **Guy:**Decimal
- **Units:**dex (logarithmic units)
- **Description:**Positive error of the stellar log g.
- **Example:**+0.07

koi_slogg_err2

- **Guy:**Decimal
- **Units:**dex
- **Description:**Negative error of the stellar log g.
- **Example:**-0.07

koi_srad

- **Guy:**Decimal
- **Units:**Solar radios (R☉)
- **Description:**Host star radius.
- **Typical range:**0.1 R☉ (small M dwarfs) to 10+ R☉ (giants)
- **Example:**1,046
- **Grades:**Sun = 1.0 R☉.Stellar radius is critical for calculating planetary size from transit depth.

koi_srad_err1

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Positive error of the stellar radius.
- **Example:**+0.053

koi_srad_err2

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Negative error of the stellar radius.
- **Example:**-0.053

10. Position and Magnitude

ra

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Right Ascension. Celestial coordinate equivalent to terrestrial longitude.
- **Range:**0.0 - 360.0
- **Example:**291.45678
- **Grades:**J2000 epoch. Kepler observed specific region in Cygnus/Lyra (RA ~ 290-300°).

dec

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Declination. Celestial coordinate equivalent to Earth's latitude.
- **Range:**-90.0 to +90.0
- **Example:**+48.12345
- **Grades:**J2000 epoch. Kepler field centered near Dec +45°.

koi_kepmag

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Kepler magnitude. The brightness of the star in the spectral band of Kepler's photometer.
- **Typical range:**9 (bright) to 17 (dim)
- **Example:**14.732
- **Grades:**Logarithmic scale: 5 magnitude difference = factor 100 in brightness. Fainter stars have higher magnitudes.

CHAPTER II. DATASET TOI (TESS OBJECTS OF INTEREST)

1. Main Identifiers

you

- **Guy:**Decimal
- **Description:**TESS Object of Interest number. Unique identifier of the candidate.
- **Format:**Decimal number (101.01, 175.02, etc.)
- **Example:**700.01
- **Grades:**Whole part = star, decimal part = planet number in the system.

tid

- **Guy:**Integer (long)
- **Description:**TESS Input Catalog ID. Identifier of the star in the TESS input catalog.
- **Typical range:**1000000 - 999999999
- **Example:**231663901
- **Grades:**Different from KepID. TIC catalog is more extensive than KIC.

tfopwg_disp

- **Guy:**Categorical text
- **Description:**Provision of the TESS Follow-up Observing Program Working Group.
- **Common values:**
  - CP: Confirmed Planet
  - PC: Planet Candidate
  - FP: False Positive
  - APC: Ambiguous Planet Candidate

  - KP: Known Planet
- **Example:**CP
- **Grades:**Classification based on follow-up observations by a global network of telescopes.

2. Astronomical Coordinates

track

- **Guy:**Text
- **Description:**Right ascension in sexagesimal format (hours:minutes:seconds).
- **Format:**HH:MM:SS.ss
- **Example:**19:23:45.67
- **Grades:**Traditional format used by astronomers. Convertible to decimal degrees.

ra

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Right ascension in decimal format.
- **Range:**0.0 - 360.0
- **Example:**290.9403
- **Grades:**Same data as rastr, different format.

decstr

- **Guy:**Text
- **Description:**Declination in sexagesimal format (degrees:minutes:seconds).
- **Format:**±DD:MM:SS.s
- **Example:**+48:14:32.1
- **Grades:**Sign indicates northern (+) or southern (-) celestial hemisphere.

dec

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Declination in decimal format.
- **Range:**-90.0 to +90.0
- **Example:**+48.2423
- **Grades:**Same data as decstr, different format.

3. Stellar Proper Motion

st_pmra

- **Guy:**Decimal

- **Units:**mas/yr (milliarcseconds per year)
- **Description:**Proper motion of the star in right ascension.
- **Typical range:**-1000 to +1000 mas/yr
- **Example:**23.456
- **Grades:**Indicates angular velocity in the sky. Nearby stars have greater proper motion.

st_pmraerr1

- **Guy:**Decimal
- **Units:**more/yr
- **Description:**Positive error of proper motion in RA.
- **Example:**+0.234

st_pmraerr2

- **Guy:**Decimal
- **Units:**more/yr
- **Description:**Negative error of proper motion in RA.
- **Example:**-0.234

st_pmralim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag. Indicates whether the st_pmra value is an upper limit rather than a measurement.
- **Values:**
  - 0: Actual measurement
  - 1: Upper limit
- **Example:**0

st_pmdec

- **Guy:**Decimal
- **Units:**more/yr
- **Description:**Proper motion of the star in declination.
- **Typical range:**-1000 to +1000 mas/yr
- **Example:**-15,789

st_pmdecerr1

- **Guy:**Decimal
- **Units:**more/yr
- **Description:**Positive error of proper motion in Dec.
- **Example:**+0.345

st_pmdecerr2

- **Guy:**Decimal
- **Units:**more/yr
- **Description:**Negative error of proper motion in Dec.
- **Example:**-0.345

st_pmdeclim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for proper movement in Dec.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

4. Time of Transition

pl_tranmid

- **Guy:**Decimal
- **Units:**BJD (Barycentric Julian Date) - 2457000
- **Description:**Time of the midpoint of the first observed transit.
- **Typical range:**1000 - 3000 (days from reference)
- **Example:**1354.6789
- **Grades:**Reference period for ephemeris calculations.

pl_tranmiderr1

- **Guy:**Decimal
- **Units:**Days
- **Description:**Positive error of the transit period.
- **Example:**+0.0012

pl_tranmiderr2

- **Guy:**Decimal
- **Units:**Days
- **Description:**Negative error of the transit period.
- **Example:**-0.0012

pl_tranmidlim

- **Guy:**Integer (0 or 1)
- **Description:**Flag indicating whether pl_tranmid is limit.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

5. Orbital Period

pl_orbper

- **Guy:**Decimal
- **Units:**Days
- **Description:**Orbital period of the planet.
- **Typical range:**0.5 - 100+ days
- **Example:**8.138
- **Grades:**TESS observes each sector for ~27 days, limiting detection of long periods in a single observation.

pl_orbpererr1

- **Guy:**Decimal
- **Units:**Days
- **Description:**Positive error of the orbital period.
- **Example:**+0.0003

pl_orbpererr2

- **Guy:**Decimal
- **Units:**Days
- **Description:**Negative orbital period error.
- **Example:**-0.0003

pl_orbperlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for orbital period.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

6. Duration of Transit

pl_trandurh

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Duration of planetary transit.
- **Typical range:**0.5 - 8 hours
- **Example:**2.45
- **Grades:**TESS has a cadence of 2 minutes (FFI mode) or 30 minutes, affecting duration accuracy.

pl_trandurherr1

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Positive duration error.
- **Example:**+0.12

pl_trandurherr2

- **Guy:**Decimal
- **Units:**Hours
- **Description:**Negative duration error.
- **Example:**-0.12

pl_trandurhlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for duration.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

7. Depth of Transit

pl_trandep

- **Guy:**Decimal
- **Units:**ppm (parts per million) or percentage
- **Description:**Depth of transit.
- **Typical range:**100 - 50,000 ppm
- **Example:**3450.0
- **Grades:**Format may vary between ppm and percentage depending on the catalog version.

pl_trandeperr1

- **Guy:**Decimal
- **Units:**ppm or percentage
- **Description:**Positive depth error.
- **Example:**+125.3

pl_trandeperr2

- **Guy:**Decimal
- **Units:**ppm or percentage
- **Description:**Negative depth error.
- **Example:**-125.3

pl_trandeplim

- **Guy:**Integer (0 or 1)
- **Description:**Depth limit flag.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

8. Planetarium Radio

pl_rade

- **Guy:**Decimal
- **Units:**Terrestrial radios (R⊕)
- **Description:**Planet Radio.
- **Typical range:**0.5 - 25 R⊕
- **Example:**2.34
- **Grades:**Derived from transit depth and stellar radius.

pl_radeerr1

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Positive error of the radius.
- **Example:**+0.23

pl_radeerr2

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Negative radius error.
- **Example:**-0.23

pl_radelim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for radius.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

9. Planetary Insolation

pl_insol

- **Guy:**Decimal
- **Units:**Overland flow (F⊕)
- **Description:**Sunstroke received by the planet.
- **Typical range:**0.1 - 10,000 F⊕
- **Example:**45.6

pl_insolerr1

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Positive insolation error.
- **Example:**+3.4

pl_insolerr2

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Negative insolation error.
- **Example:**-3.4

pl_insollim

- **Guy:**Integer (0 or 1)
- **Description:**Flag limit for heat stroke.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

10. Equilibrium Temperature

pl_eqt

- **Guy:**Decimal
- **Units:**Kelvin (K)
- * * Description:** Equilibrium temperature of the planet.
- **Typical range:**200 - 3000 K
- **Example:**856
- **Grades:**Assumes perfect heat redistribution and albedo = 0.

pl_eqterr1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive error of the equilibrium temperature.
- **Example:**+34

pl_eqterr2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative error of the equilibrium temperature.
- **Example:**-34

pl_eqtlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for temperature.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

11. TESS Magnitude

st_tmag

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**TESS magnitude of the star. Brightness in the TESS spectral band (600-1000 nm).
- **Typical range:**4 - 16
- **Example:**11.234
- **Grades:**TESS prioritizes bright stars (Tmag < 12) for better follow-up data.

st_tmagerr1

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Positive error of the TESS magnitude.
- **Example:**+0.012

st_tmagerr2

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Negative error of the TESS magnitude.
- **Example:**-0.012

st_tmaglim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for magnitude.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

12. Stellar Distance

st_dist

- **Guy:**Decimal

- **Units:**Parsecs (pc)
- **Description:**Distance to the host star.
- **Typical range:**10 - 500 pc
- **Example:**127.8
- **Grades:**1 parsec = 3.26 light-years. TESS focuses on nearby stars to facilitate tracking.

st_disterr1

- **Guy:**Decimal
- **Units:**Parsecs
- **Description:**Positive distance error.
- **Example:**+5.4

st_disterr2

- **Guy:**Decimal
- **Units:**Parsecs
- **Description:**Negative distance error.
- **Example:**-5.4

st_distlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for distance.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

13. Stellar Temperature

st_teff

- **Guy:**Decimal
- **Units:**Kelvin (K)
- **Description:**Effective temperature of the star.
- **Typical range:**2500 - 10,000 K
- **Example:**5456
- **Grades:**Determines spectral type (M, K, G, F, A).

st_tefferr1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive stellar temperature error.
- **Example:**+87

st_tefferr2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative stellar temperature error.
- **Example:**-87

st_tefflim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for temperature.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

14. Stellar Surface Gravity

st_logg

- **Guy:**Decimal
- **Units:**$\log_{10}(cm/s^2)$
- **Description:**Logarithm of stellar surface gravity.
- **Typical range:**3.5 - 5.0
- **Example:**4.52

st_loggerr1

- **Guy:**Decimal
- **Units:**dex
- **Description:**Positive log g error.
- **Example:**+0.09

st_loggerr2

- **Guy:**Decimal
- **Units:**dex
- **Description:**Negative log g error.
- **Example:**-0.09

st_logglim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for log g.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

15. Radio Estelar

st_rad

- **Guy:**Decimal
- **Units:**Solar radios (R☉)
- **Description:**Host star radius.
- **Typical range:**0.1 - 3 R☉
- **Example:**0.987

st_raderr1

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Positive error of the stellar radius.
- **Example:**+0.045

st_raderr2

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Negative error of the stellar radius.
- **Example:**-0.045

st_radlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for radius.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

16. Metadata

toi_created

- **Guy:**Date
- **Format:**YYYY-MM-DD
- **Description:**Date the TOI was created in the catalog.
- **Example:**2019-03-15
- **Grades:**Indicates when the candidate was identified and classified.

rowupdate

- **Guy:**Date/Timestamp
- **Format:**YYYY-MM-DD HH:MM:SS
- **Description:**Date and time of the last log update.
- **Example:**2023-08-22 14:35:22
- **Grades:**The TOI catalog is regularly updated with new data and provisions.

CHAPTER III. DATASET K2 (K2 PLANETS AND CANDIDATES)

1. Nomenclature and Classification

pl_name

- **Guy:**Text
- **Description:**Official planet name confirmed.
- **Format:**K2-### + letter (K2-18b)
- **Example:**K2-18b
- **Grades:**Only present for confirmed planets. Candidates use EPIC nomenclature.

hostname

- **Guy:**Text
- **Description:**Name of the host star.
- **Format:**K2-### or EPIC ########
- **Example:**K2-18
- **Grades:**EPIC = Ecliptic Plane Input Catalog used by K2.

default_flag

- **Guy:**Integer (0 or 1)
- **Description:**Flag indicating whether this is the default parameter when multiple measurements exist.
- **Values:**
  - 1: Recommended/Default Parameters
  - 0: Alternative measurements
- **Example:**1
- **Grades:**Useful when multiple studies report different values.

disposition

- **Guy:**Categorical text
- **Description:**Planet confirmation status.
- **Common values:**
  - CONFIRMED: Planet confirmed
  - CANDIDATE: Candidate not yet confirmed
  - FALSE POSITIVE: Discarded as a planet
- **Example:**CONFIRMED

disp_refname

- **Guy:**Text
- **Description:**Bibliographic reference (paper) that established the provision.
- **Format:**Author et al. year
- **Example:**Crossfield et al. 2016
- **Grades:**Link to a scientific publication that confirmed or ruled out the planet.

sy_snum

- **Guy:**Whole
- **Description:**Number of stars in the system.
- **Typical values:**1 (simple systems), 2 (binary), 3+ (multiple)
- **Example:**1
- **Grades:**Most are single star systems.

sy_pnum

- **Guy:**Whole
- **Description:**Number of known planets in the system.
- **Range:**1 - 7
- **Example:**2
- **Grades:**K2 discovered several multi-planetary systems.

2. Discovery Method

discovery method

- **Guy:**Categorical text
- **Description:**Method used to detect the planet.
- **Common values:**
    - Transit: Transit method
    - Radial Velocity: Radial velocity
    - Microlensing: Gravitational microlensing
- **Example:**Transit
- **Grades:**K2 mostly uses transit, but some planets confirmed by RV.

disc_year

- **Guy:**Whole
- **Description:**Year of discovery/publication.
- **Typical range:**2014 - 2020
- **Example:**2017

disc_facility

- **Guy:**Text
- **Description:**Facility/telescope that discovered the planet.
- **Common values:**
    - K2: K2 Mission
    - Kepler: Original Kepler Mission
    - Ground: Ground-based telescopes
- **Example:**K2

soltype

- **Guy:**Text
- **Description:**Type of orbital solution.
- **Values:**
    - Published Confirmed: Planet parameters confirmed
    - Controversial: Controversial detection
- **Example:**Published Confirmed

pl_controv_flag

- **Guy:**Integer (0 or 1)
- **Description:**Flag indicating whether the planet is controversial.
- **Values:**
    - 0: Not controversial
    - 1: Disputed or uncertain detection
- **Example:**0

pl_refname

- **Guy:**Text
- **Description:**Main bibliographic reference of the planet.
- **Format:**Author et al. year
- **Example:**Montet et al. 2015

3. Orbital Parameters

pl_orbper

- **Guy:**Decimal
- **Units:**Days
- **Description:**Orbital period of the planet.
- **Typical range:**0.5 - 100+ days
- **Example:**32.9396

pl_orbpererr1

- **Guy:**Decimal

- **Units:**Days
- **Description:**Positive period error.
- **Example:**+0.0012

pl_orbpererr2

- **Guy:**Decimal
- **Units:**Days
- **Description:**Negative period error.
- **Example:**-0.0012

pl_orbperlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for period.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

pl_orbsmax

- **Guy:**Decimal
- **Units:**AU (Astronomical Units)
- **Description:**Semi-major axis of the orbit. Average planet-star distance.
- **Typical range:**0.01 - 5 AU
- **Example:**0.1429
- **Grades:**1 AU = Earth-Sun distance. Mercury = 0.39 AU.

pl_orbsmaxerr1

- **Guy:**Decimal
- **Units:**AU
- **Description:**Positive error of the semi-major axis.
- **Example:**+0.0023

pl_orbsmaxerr2

- **Guy:**Decimal
- **Units:**AU
- **Description:**Negative error of the semi-major axis.
- **Example:**-0.0023

pl_orbsmaxlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for semi-major axis.
- **Values:**0 = measurement, 1 = limit

- **Example:**0

4. Planetarium Radio

pl_rade

- **Guy:**Decimal
- **Units:**Terrestrial radios (R⊕)
- **Description:**Planetary radius in terrestrial radios.
- **Typical range:**0.5 - 25 R⊕
- **Example:**2.24

pl_radeerr1

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Positive error of the radius.
- **Example:**+0.13

pl_radeerr2

- **Guy:**Decimal
- **Units:**Terrestrial radios
- **Description:**Negative radius error.
- **Example:**-0.13

pl_radelim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for radius in R⊕.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

pl_radj

- **Guy:**Decimal
- **Units:**Jovian radios (RJ)
- **Description:**Planet radius in Jupiter radii.
- **Typical range:**0.04 - 2 RJ
- **Example:**0.200
- **Grades:**1 RJ ≈ 11.2 R⊕.Useful for gas giants.

pl_radjerr1

- **Guy:**Decimal
- **Units:**Jovian radios

- **Description:**Positive error of the Jovian radius.
- **Example:**+0.012

pl_radjerr2

- **Guy:**Decimal
- **Units:**Jovian radios
- **Description:**Negative error of the Jovian radius.
- **Example:**-0.012

pl_radjlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for radio in RJ.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

5. Planetary Mass

pl_bmasse

- **Guy:**Decimal
- **Units:**Land masses (M⊕)
- **Description:**Mass of the planet in land masses.
- **Typical range:**0.5 - 1000 M⊕
- **Example:**8.63
- **Grades:**Requires radial velocity measurements, not available with transits alone.

pl_bmasseerr1

- **Guy:**Decimal
- **Units:**Land masses
- **Description:**Positive mass error.
- **Example:**+1.35

pl_bmasseerr2

- **Guy:**Decimal
- **Units:**Land masses
- **Description:**Negative mass error.
- **Example:**-1.35

pl_bmasselim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for mass in M⊕.

- **Values:** 0 = measurement, 1 = limit
- **Example:** 0

pl_bmassj

- **Guy:** Decimal
- **Units:** Jovian masses (JM)
- **Description:** Planet mass in Jupiter masses.
- **Typical range:** 0.001 - 3 MJ
- **Example:** 0.0272
- **Grades:** 1 MJ ≈ 318 M⊕.

pl_bmassjerr1

- **Guy:** Decimal
- **Units:** Jovian masses
- **Description:** Positive Jovian mass error.
- **Example:** +0.0043

pl_bmassjerr2

- **Guy:** Decimal
- **Units:** Jovian masses
- **Description:** Negative Jovian mass error.
- **Example:** -0.0043

pl_bmassjlim

- **Guy:** Integer (0 or 1)
- **Description:** Limit flag for mass in MJ.
- **Values:** 0 = measurement, 1 = limit
- **Example:** 0

pl_bmassprov

- **Guy:** Text
- **Description:** Origin of mass measurement.
- **Common values:**
  - Msin(i): Minimum RV mass (without knowing inclination)
  - Mass: True mass
  - Msini: Alternative notation
- **Example:** Mass
- **Grades:** Transits provide tilt, allowing true mass.

6. Orbital Eccentricity

pl_orbeccen

- **Guy:**Decimal
- **Description:**Orbital eccentricity. How elliptical the orbit is.
- **Range:**0.0 (circular) to ~0.9 (very elliptical)
- **Example:**0.041
- **Grades:**Earth = 0.017. Circular orbits are e ≈ 0.

pl_orbeccenerr1

- **Guy:**Decimal
- **Description:**Positive eccentricity error.
- **Example:**+0.023

pl_orbeccenerr2

- **Guy:**Decimal
- **Description:**Negative eccentricity error.
- **Example:**-0.023

pl_orbeccenlim

- **Guy:**Integer (0 or 1)
- **Description:**Eccentricity limit flag.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

7. Sunstroke and Temperature

pl_insol

- **Guy:**Decimal
- **Units:**Overland flow (F⊕)
- **Description:**Sunstroke received by the planet.
- **Typical range:**0.1 - 10,000 F⊕
- **Example:**1.22

pl_insolerr1

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Positive insolation error.
- **Example:**+0.08

pl_insolerr2

- **Guy:**Decimal
- **Units:**Overland flow
- **Description:**Negative insolation error.
- **Example:**-0.08

pl_insollim

- **Guy:**Integer (0 or 1)
- **Description:**Flag limit for heat stroke.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

pl_eqt

- **Guy:**Decimal
- **Units:**Kelvin (K)
- **Description:**Planetary equilibrium temperature.
- **Typical range:**200 - 3000 K
- **Example:**265

pl_eqterr1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive temperature error.
- **Example:**+7

pl_eqterr2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative temperature error.
- **Example:**-7

pl_eqtlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for temperature.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

8. Transit Timing Variations

ttv_flag

- **Guy:**Integer (0 or 1)

- **Description:**Flag indicating whether transit time variations (TTV) were detected.
- **Values:**
  - 0: No TTVs were detected
  - 1: TTVs detected
- **Example:**0
- **Grades:**TTVs indicate gravitational disturbances from other planets in the system.

9. Properties of the Host Star

st_refname

- **Guy:**Text
- **Description:**Bibliographic reference for stellar parameters.
- **Format:**Author et al. year
- **Example:**Huber et al. 2016

st_spectype

- **Guy:**Text
- **Description:**Spectral type of the star.
- **Format:**Letter + number (M3V, G2V, K5V, etc.)
- **Example:**M2.5V
- **Grades:**OBAFGKM (hot to cold). V = main sequence.

st_teff

- **Guy:**Decimal
- **Units:**Kelvin (K)
- **Description:**Stellar effective temperature.
- **Typical range:**2500 - 10,000 K
- **Example:**3503

st_tefferr1

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Positive stellar temperature error.
- **Example:**+73

st_tefferr2

- **Guy:**Decimal
- **Units:**Kelvin
- **Description:**Negative stellar temperature error.
- **Example:**-73

st_tefflim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for stellar temperature.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

st_rad

- **Guy:**Decimal
- **Units:**Solar radios (R☉)
- **Description:**Stellar radio.
- **Typical range:**0.1 - 10 R☉
- **Example:**0.411

st_raderr1

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Positive error of the stellar radius.
- **Example:**+0.017

st_raderr2

- **Guy:**Decimal
- **Units:**Solar radios
- **Description:**Negative error of the stellar radius.
- **Example:**-0.017

st_radlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for stellar radius.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

st_mass

- **Guy:**Decimal
- **Units:**Solar masses (M☉)
- **Description:**Mass of the star.
- **Typical range:**0.1 - 3 M☉
- **Example:**0.412

st_masserr1

- **Guy:**Decimal
- **Units:**Solar masses
- **Description:**Positive stellar mass error.
- **Example:**+0.023

st_masserr2

- **Guy:**Decimal
- **Units:**Solar masses
- **Description:**Negative stellar mass error.
- **Example:**-0.023

st_masslim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for stellar mass.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

10. Stellar Metallicity

st_met

- **Guy:**Decimal
- **Units:**dex (logarithmic scale)
- **Description:**Stellar metallicity [Fe/H]. Abundance of elements heavier than helium relative to the Sun.
- **Typical range:**-1.0 (metal-poor) to +0.5 (metal-rich)
- **Example:**-0.12
- **Grades:**Sun = 0.0. Metallicity affects planetary formation.

st_meterr1

- **Guy:**Decimal
- **Units:**dex
- **Description:**Positive metallicity error.
- **Example:**+0.08

st_meterr2

- **Guy:**Decimal
- **Units:**dex
- **Description:**Negative metallicity error.
- **Example:**-0.08

st_metlim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for metallicity.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

st_metratio

- **Guy:**Text
- **Description:**Specific metallicity ratio used (example: [Fe/H], [M/H]).
- **Example:**[Fe/H]
- **Grades:**[[Fe/H] = iron abundance. [M/H] = all metals.

st_logg

- **Guy:**Decimal
- **Units:**$\log_{10}(cm/s^2)$
- **Description:**Logarithm of stellar surface gravity.
- **Typical range:**3.5 - 5.0
- **Example:**4.82

st_loggerr1

- **Guy:**Decimal
- **Units:**dex
- **Description:**Positive log g error.
- **Example:**+0.04

st_loggerr2

- **Guy:**Decimal
- **Units:**dex
- **Description:**Negative log g error.
- **Example:**-0.04

st_logglim

- **Guy:**Integer (0 or 1)
- **Description:**Limit flag for log g.
- **Values:**0 = measurement, 1 = limit
- **Example:**0

11. System Properties

sy_refname

- **Guy:**Text

- **Description:**Bibliographic reference for system properties.
- **Format:**Author et al. year
- **Example:**Montet et al. 2015

ra

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Right ascension of the system.
- **Range:**0.0 - 360.0
- **Example:**186.456

track

- **Guy:**Text
- **Description:**Right ascension in sexagesimal format.
- **Format:**HH:MM:SS.ss
- **Example:**12:25:49.5

dec

- **Guy:**Decimal
- **Units:**Decimal degrees
- **Description:**Decline of the system.
- **Range:**-90.0 to +90.0
- **Example:**+7.823

decstr

- **Guy:**Text
- **Description:**Declension in sexagesimal format.
- **Format:**±DD:MM:SS.s
- **Example:**+07:49:23.4

sy_dist

- **Guy:**Decimal
- **Units:**Parsecs (pc)
- **Description:**Distance to the system.
- **Typical range:**10 - 1000 pc
- **Example:**38.12

sy_disterr1

- **Guy:**Decimal
- **Units:**Parsecs

- **Description:**Positive distance error.
- **Example:**+1.23

sy_disterr2

- **Guy:**Decimal
- **Units:**Parsecs
- **Description:**Negative distance error.
- **Example:**-1.23

12. Photometric Magnitudes

sy_vmag

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Magnitude V (visual) of the system.
- **Typical range:**8 - 18
- **Example:**13.467

sy_vmagerr1

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Positive error of magnitude V.
- **Example:**+0.012

sy_vmagerr2

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Negative error of magnitude V.
- **Example:**-0.012

sy_kmag

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**K (near infrared) magnitude of the system.
- **Typical range:**7 - 16
- **Example:**9.234

sy_kmagerr1

- **Guy:**Decimal
- **Units:**Magnitude

- **Description:**Positive error of magnitude K.
- **Example:**+0.023

sy_kmagerr2

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Negative error of magnitude K.
- **Example:**-0.023

sy_gaiamag

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Gaia Magnitude (band

Data Dictionary – Datasets NASA exoplanets

(Continuation)

G) of the Gaia satellite.

- **Typical range:**8 - 18
- **Example:**12.789
- **Grades:**Gaia DR2/DR3 provides precise photometry and astrometry.

sy_gaiamagerr1

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Gaia magnitude positive error.
- **Example:**+0.008

sy_gaiamagerr2

- **Guy:**Decimal
- **Units:**Magnitude
- **Description:**Gaia magnitude negative error.
- **Example:**-0.008

13. Metadata and Dates

rowupdate

- **Guy:**Date/Timestamp
- **Format:**YYYY-MM-DD HH:MM:SS
- **Description:**Date and time the record was last updated.
- **Example:**2023-05-15 10:23:45
- **Grades:**Indicates when the entry was last modified.

pl_pubdate

- **Guy:**Date
- **Format:**YYYY-MM-DD
- **Description:**Date of publication of the planet's discovery.
- **Example:**2016-07-18
- **Grades:**Date of the scientific paper that announced the planet.

releasedate

- **Guy:**Date
- **Format:**YYYY-MM-DD
- **Description:**Date of release of the data in the NASA exoplanet archive.
- **Example:**2016-08-01
- **Grades:**May differ from pl_pubdate due to data curation processes.

CHAPTER IV. INTERPRETATION OF FLAGS AND LIMITS

1. Meaning of Limit Flags

Limit flags (columns ending in "lim") indicate the nature of the measurement:

Value 0 (Actual Measurement): The reported value is a direct measurement with uncertainty quantified by the err1 and err2 fields.

Value 1 (Upper Limit): The reported value is an upper limit. The true parameter is less than or equal to the stated value, but it could not be measured accurately.

Example of Interpretation:

```
None
● st_pmra = 50.5
● st_pmraerr1 = +2.3
```

- st_pmraerr2 = -2.3
- st_pmralim = 0
-

Interpretation: Measured proper motion = 50.5 ± 2.3 mas/yr

None

- pl_bmasse = 15.0
- pl_bmasselim = 1
-

Interpretation: Planetary mass ≤ 15.0 M⊕ (upper limit)

2. Null and Missing Values

NULL or empty values: They indicate that the parameter has not been measured or is not applicable for that object.

Common reasons for missing values:

For planets detected only by transit:

- Planetary mass (requires radial velocity)
- Orbital eccentricity (requires multiple methods)

For unconfirmed candidates:

- Official name of the planet (pl_name, kepler_name)
- Some refined stellar parameters

For objects without spectroscopic tracking:

- Detailed stellar metallicity
- Precise spectral type

3. Asymmetric Errors

Many columns have err1 (positive error) and err2 (negative error) that may differ in magnitude:

Reasons for asymmetry:

- Non-Gaussian probability distributions
- Physical limits (example: eccentricity cannot be negative)
- Uncertainties arising from other parameters with asymmetric errors

Standard notation:

```
None
● Value = X +err1 -|err2|
●
● Example:
● koi_prad = 2.34 +0.18 -0.15 R⊕
●

Means: Radius = 2.34 with range [2.19, 2.52] R⊕
```

CHAPTER V. CONSIDERATIONS FOR USE IN MACHINE LEARNING

1. Recommended Preprocessing

Handling Missing Values:

Strategy 1: Elimination

- Delete rows with >30% missing values
- Remove columns with >50% missing values

Strategy 2: Imputation

- Mean/median for continuous variables
- Mode for categorical variables
- Additional flags indicating imputed value

Treatment of Limit Flags:

Option 1: Treat as regular values

- Use the numeric value directly
- You will lose information about uncertainty

Option 2: Create indicator variable

- Additional column: 1 if limit, 0 if measurement
- Maintains information on data quality

2. Critical Columns per Dataset

For KOI (disposition prediction):

Most informative variables:

- koi_score (critical)
- koi_fpflag_* (all flags)
- koi_period
- koi_depth
- koi_duration
- koi_prad
- koi_model_snr
- koi_steff
- koi_srad

For TOI (disposition prediction):

Most informative variables:

- pl_trandep
- pl_rade
- pl_orbper
- st_teff
- st_rad
- st_dist
- st_tmag (data quality indicator)

For K2 (if target includes provision):

Most informative variables:

- pl_rade or pl_radj
- pl_orbper
- pl_insol
- st_teff

- st_rad
- st_mass
- sy_dist

3. Useful Derived Variables

Planetary Density:

```
None
```

- ρ_planet∝mass / radius^3
-
- Useful to distinguish:
- - Gas giants (low density)
- - Rocky (high density)

- Neptunes (intermediate density)

Reason Radio Planeta/Radio Estrella:

```
None
```

- Rp/Rs∝ √transit_depth
-

Directly related to observable depth

Normalized Equilibrium Temperature:

```
None
```

- Teq / Teff_estrella
-

Indicates relative thermal environment

Habitable Zone (binary flag):

```
None
```

- 1 if 0.35 < heatstroke < 1.7 F⊕

0 otherwise

4. Important Correlations

Highly Correlated (avoid redundancy):

- pl_rade and pl_radj (same parameter, different units)
- pl_bmasse and pl_bmassj (same parameter, different units)
- ra/rastr and dec/decstr (different format, same data)

Physically Related:

- koi_depth and koi_prad (depth∝radius²)
- koi_insol and koi_teq (temperature∝√heatstroke)
- koi_period and pl_orbsmax (Kepler's third law)

5. Scaling and Normalization

Variables with Extreme Ranges:

Naturally logarithmic:

- Orbital period (0.5 - 700 days): use log10
- Heatstroke (0.1 - 10,000 F)⊕):use log10
- Stellar distance (10 - 1000 pc): use log10

Approximate normal distribution:

- Temperatures (K): z-score normalization
- Radios (R⊕,R☉):z-score normalization
- Stellar metallicity (dex): already on a logarithmic scale

Recommended Strategy:

1. Identify the distribution of each variable
2. Apply logarithmic transformation if there are orders of magnitude of range
3. Standardize (z-score) after transformations

CHAPTER VI. DATA VALIDATION

1. Physically Plausible Ranges

Health Checks:

Orbital Period:

- Physical minimum: ~0.1 days (tidal break limit)
- Maximum observable: depends on mission duration

Planetarium Radio:

- Minimum: ~0.3 R⊕ (Mercury-like)
- Maximum: ~2 RJ (limit before being a brown dwarf)

Equilibrium Temperature:

- Minimum: ~50 K (outer limits)
- Maximum: ~3500 K (ultra-hot Jupiters)

Transit Depth:

- Minimum detectable: ~20 ppm (Kepler limit)
- Physical maximum: ~3% (Jupiter transiting M dwarf)

2. Internal Consistency

Cross-checks:

Depth vs Radius:

```None
● depth (ppm) ≈ 1,000,000 × (Rplanet/Star)²
●

Verify that values   are consistent within errors
```

Sunstroke vs Temperature:

```None
● Teq ≈ Teff × √(Star/2a) × (1-A)^0.25
●
```

```
where A = albedo (~0 for dark bodies)
```

Period vs. Semimajor axis (Kepler's 3rd Law):

```
None
● P²∝a³ / Mestrella
●

Check consistency if both are available
```

3. Outlier Detection

Potentially Problematic Outliers:

Very low SNR (<7):

- They may be spurious detections
- Consider removing or marking as low confidence

Very large errors (>50% of the value):

- Very uncertain measurement
- Evaluate whether to include in training

Multiple flags active:

- koi_fpflag_nt + koi_fpflag_ss + koi_fpflag_co all = 1
- Almost certainly false positive

Scientifically Interesting Outliers:

Ultra-short planets (P < 1 day):

- Rare but real
- Do not delete automatically

Giants in wide orbits (P > 100 days):

- Difficult to detect but they exist
- Check with multiple sources

CHAPTER VII. REFERENCES AND RESOURCES

1. Official Documentation

NASA Exoplanet Archive:https://exoplanetarchive.ipac.caltech.edu/docs/

Contains:

- Detailed descriptions of each column
- Methodologies for calculating derived parameters
- Updates and changes to catalogs
- API usage tutorials

Kepler Mission:https://www.nasa.gov/mission_pages/kepler/

Documentation on:

- Mission design
- Data processing pipeline
- Manual vetting procedures

TESS Mission:https://tess.mit.edu/

Information about:

- Observation strategy
- Observation cadences and modes
- TESS magnitude system

K2 Mission:https://www.nasa.gov/mission_pages/kepler/k2/

Details about:

- Observation campaigns
- Differences with the original Kepler mission
- K2 data characteristics

2. Glossary of Acronyms

BJD: Barycentric Julian Date – Time referenced to the barycenter of the Solar System

BKJD: Barycentric Kepler Julian Date - BJD - 2454833

EPIC: Ecliptic Plane Input Catalog - Input catalog for K2

KIC: Kepler Input Catalog - Kepler input catalog

KOI: Kepler Object of Interest - Kepler Object of Interest

TIC:TESS Input Catalog - Input catalog for TESS

TOI:TESS Object of Interest - TESS Object of Interest

TCE: Threshold Crossing Event - Event that crosses the detection threshold

SNR: Signal-to-Noise Ratio

TTV: Transit Timing Variation - Variation in transit time

ppm: Parts per million - Parts per million

mas/yr: Milliarcseconds per year - Milliarcseconds per year

AU: Astronomical Unit - Astronomical Unit (Earth-Sun distance)

R⊕: Earth radius

RJ: Jovian radius (Jupiter radius)

M⊕: Land mass

MJ: Jupiter mass

R☉: Solar radio

M☉: Solar mass

F⊕: Overland flow

APPENDIX: QUICK REFERENCE TABLES

Table 1: Unit Conversions

None

- Radios:
- 1 RJ = 11.2 R⊕ =0.1 R☉
- 1 R⊕ =6,371 km

-
- Masses:
- 1 MJ = 318 M⊕ =0.001 M☉
- 1 M⊕ =5.972 × 10^{24} kg
-
- Distances:
- 1 AU = 1.496 × 10^{8}km
- 1 parsec = 3.26 light years = 206,265 AU
-
- Temperatures:
- 0°C = 273.15 K

Tsol = 5778 K

Table 2: Classification by Radio

None

- Radio (R⊕) | Type
- - - - - - - - - - - - - - - - |------------------------
- <1.25 | Sub-Earth / Terrestrial | Super-
- 1.25 - 2.0   Earth
- 2.0 - 4.0 | Mini-Neptune
- 4.0 - 8.0 | Sub-Neptune / Neptune |
- 8.0 - 14.0   Sub-Jupiter

> 14.0 | Jupiter or greater

Table 3: Habitable Zone (Conservative)

None

- Insolation | Classification
- - - - - - - - - - - - - - - - |------------------------
- > 1.7 F⊕ | Too hot
- 0.35 - 1.7 F⊕ |Habitable Zone

```
< 0.35 F⊕          | Too cold
```

Table 4: Stellar Spectral Types

```
None

● Type | Teff (K) | Color                    | Examples
● - - - - - |----------|------------|--------------
● EITHER    | 30,000+      | Blue                | Very rare
● B         | 10,000       | Blue-White | Rigel
● TO        | 7,500        | White               | Vega, Sirius
● F         | 6,000        | Yellow-W | Procyon |
● G         | 5,500        Yellow                | Sun (G2V)
● K         | 4,500        | Orange             | Arcturus

M       | 3,000          | Red                | Proxima Centauri
```

END OF DATA DICTIONARY

This document provides the technical foundation needed to understand, process, and analyze NASA's exoplanet datasets. For up-to-date information on changes to the data structures, please always refer to the official NASA Exoplanet Archive documentation.