# Project Documentation: A World Away - Hunting for Exoplanets with AI

## 1. Introduction

This document outlines the project developed in response to the NASA Space Apps Challenge 2025, titled "A World Away: Hunting for Exoplanets with AI." The primary goal of this project is to leverage Artificial Intelligence and Machine Learning (AI/ML) to automate the detection and classification of exoplanets using publicly available NASA datasets. This initiative aims to accelerate the discovery process of planets outside our solar system, moving beyond traditional manual analysis methods.

## 2. Challenge Overview

The "A World Away: Hunting for Exoplanets with AI" challenge addresses the growing need for automated solutions in exoplanetary science. Historically, the identification of exoplanets, particularly those detected via the transit method, has been a labor-intensive process. Missions such as Kepler, K2, and TESS have generated vast quantities of data, leading to the discovery of thousands of exoplanets. However, the sheer volume of this data necessitates advanced computational approaches to efficiently identify new candidates and confirm existing ones.

### 2.1. Problem Statement

The core problem is the manual and time-consuming nature of exoplanet identification from large astronomical datasets. While human expertise is invaluable, the scale of data from modern space telescopes overwhelms manual processing capabilities, potentially delaying discoveries and limiting the scope of research.

### 2.2. Challenge Objectives

The challenge mandates the creation of an AI/ML model trained on NASA's open-source exoplanet datasets. Beyond the model's predictive capabilities, a critical component is the development of a user-friendly web interface. This interface should enable users to interact with the model, analyze new data, and potentially contribute to its ongoing improvement. The project encourages flexibility in technology choices, allowing participants to utilize any open-source programming languages, machine learning libraries, or software solutions.

## 3. Background on Exoplanet Detection

Exoplanetary science is a rapidly evolving field, driven by dedicated space missions designed to discover and characterize planets beyond our solar system. The primary method relevant to this challenge is the **transit method**.

### 3.1. The Transit Method

The transit method involves observing a slight, periodic decrease in the brightness of a star. This dimming occurs when an exoplanet passes directly in front of its host star from the perspective of an observer (e.g., a space telescope). The duration and depth of this dip in light provide crucial information about the exoplanet's size, orbital period, and other characteristics. This method has been highly successful, leading to the discovery of the majority of known exoplanets.

### 3.2. Key Missions and Data Sources

Several NASA missions have been instrumental in collecting the data used for exoplanet detection:

* **Kepler Space Telescope:** Launched in 2009, Kepler was a pioneering mission that observed a single region of the sky for nearly a decade. It provided an unprecedented dataset for exoplanet transits, leading to thousands of discoveries.
* **K2 Mission:** Following Kepler's initial mission, the K2 mission utilized the same hardware but surveyed different fields of view across the ecliptic plane. It continued to collect valuable transit data, albeit with a different observing strategy.
* **Transiting Exoplanet Survey Satellite (TESS):** Launched in 2018, TESS is Kepler's successor, designed to perform an all-sky survey for transiting exoplanets. TESS focuses on brighter, nearby stars, making it easier for ground-based telescopes to follow up on discoveries and characterize the exoplanets.

These missions have generated extensive datasets, which include not only confirmed exoplanets but also planetary candidates and false positives. This labeled data is crucial for training supervised machine learning models.

## 4. Project Approach and Methodology

Our project focuses on developing a robust AI/ML pipeline for exoplanet identification, complemented by an intuitive web interface.

### 4.1. Data Acquisition and Preprocessing

The project will utilize the publicly available datasets from NASA, specifically:

* **Kepler Objects of Interest (KOI):** This dataset contains comprehensive information on confirmed exoplanets, candidates, and false positives from the Kepler mission. Key variables include orbital period, transit duration, and planetary radius.
* **TESS Objects of Interest (TOI):** Similar to KOI, this dataset provides classifications (confirmed exoplanets, planetary candidates, false positives, ambiguous planetary candidates, known planets) from the TESS mission.
* **K2 Planets and Candidates:** This dataset offers classifications from the K2 mission.

Preprocessing steps will involve handling missing values, normalizing features, and potentially engineering new features from the raw data to enhance model performance. The choice of which variables to include or exclude will be critical, as highlighted in the challenge description.

### 4.2. Machine Learning Model Development

The core of the project involves training an AI/ML model. Based on the challenge resources, ensemble-based machine learning algorithms have shown high accuracy in exoplanet identification. We will explore models such as Random Forests, Gradient Boosting Machines (e.g., XGBoost, LightGBM), or potentially deep learning architectures like Convolutional Neural Networks (CNNs) if the data format (e.g., light curves) allows for it. The model will be trained to classify data points into categories such as 'confirmed exoplanet,' 'planetary candidate,' or 'false positive.'

### 4.3. Web Interface Design and Implementation

A user-friendly web interface is a mandatory component. This interface will serve several functions:

* **Data Upload/Input:** Allow users to upload new astronomical data (e.g., light curves or pre-processed features) or manually input parameters for analysis.
* **Exoplanet Prediction:** Display the model's predictions for new data, indicating whether a data point is classified as an exoplanet, candidate, or false positive.
* **Model Performance Metrics:** Present statistics on the model's accuracy, precision, recall, and F1-score, providing transparency into its performance.
* **Hyperparameter Tuning (Optional):** Potentially allow advanced users to tweak certain model hyperparameters through the interface to observe their impact on predictions.
* **Visualization:** Visualize light curves, transit events, and other relevant data to help users understand the model's decisions and the characteristics of potential exoplanets.

## 5. Potential Considerations and Future Enhancements

### 5.1. Target Audience Focus

The project can be tailored for different audiences:

* **Researchers:** Providing tools for rapid classification of new data, advanced visualization, and integration with existing astronomical workflows.
* **Novices/Educators:** Offering an accessible platform to explore exoplanet data, understand the transit method, and learn about AI/ML applications in space science.

### 5.2. Continuous Learning and Model Updates

An advanced feature could involve incorporating user-provided data to continuously update and improve the model. This could be implemented through active learning strategies or periodic retraining with newly validated data.

### 5.3. Integration with Other Data Sources

While the primary focus is on Kepler, K2, and TESS data, future enhancements could involve integrating data from other missions or ground-based observatories, such as NEOSSat or even data relevant to JWST observations, to broaden the scope of exoplanet detection.

## 6. Conclusion

This project aims to deliver an innovative AI/ML solution for exoplanet detection, making the process more efficient and accessible. By combining robust machine learning models with an interactive web interface, we seek to contribute to the ongoing quest for understanding our universe and discovering new worlds. The project aligns with NASA's vision of open science and leveraging cutting-edge technology for scientific exploration.

## 7. Frequently Asked Questions (FAQ)

**Q1: What is the main goal of this project?**
A1: The main goal is to use Artificial Intelligence and Machine Learning (AI/ML) to automatically detect and classify exoplanets from NASA's open-source datasets, moving away from manual identification methods.

**Q2: Which NASA missions' data are used in this project?**
A2: This project primarily uses data from the Kepler, K2, and TESS missions, which are publicly available.

**Q3: What is the 'transit method' of exoplanet detection?**
A3: The transit method detects exoplanets by observing a slight, periodic decrease in a star's brightness when an exoplanet passes in front of it. This dimming provides information about the exoplanet's size and orbit.

**Q4: What kind of AI/ML model is being developed?**
A4: We are developing an AI/ML model, potentially using ensemble-based algorithms or deep learning, to classify data points as 'confirmed exoplanet,' 'planetary candidate,' or 'false positive.'

**Q5: Is there a user interface for this project?**
A5: Yes, a user-friendly web interface is a critical component. It will allow users to upload new data, view predictions, and potentially interact with model parameters.

**Q6: Can users contribute new data to improve the model?**
A6: While not a mandatory feature, the project considers the possibility of allowing users to upload new data to continuously update and improve the model through active learning strategies.

**Q7: What are the potential audiences for this project?**
A7: The project can be tailored for both researchers who need rapid classification tools and novices/educators interested in exploring exoplanet data and learning about AI/ML in space science.

**Q8: What programming languages or libraries are used?**
A8: The challenge encourages flexibility, allowing the use of any open-source programming language, machine learning libraries, or software solutions that fit the project well.

**Q9: How does this project help in exoplanet discovery?**
A9: By automating the detection process, the project makes exoplanet identification more efficient and accessible, helping to uncover new exoplanets hidden within vast astronomical datasets.

**Q10: Where can I find the datasets used in this project?**
A10: The datasets (Kepler Objects of Interest, TESS Objects of Interest, K2 Planets and Candidates) are publicly available through NASA's resources, as linked in the challenge description.