Executive Summary: Exoplanet Hunter AI

This project implements an end-to-end machine learning system to automate the classification of exoplanet candidates from NASA's Kepler Space Telescope data. The solution consists of a Python-based data processing and model training pipeline, a RESTful API built with Flask to serve predictions, and a dynamic web interface for user interaction. The system effectively transforms a time-intensive manual analysis task into a rapid, automated workflow, demonstrating a practical application of AI in accelerating scientific research.

Problem Statement

The exponential growth of astrophysical data from sky survey missions like Kepler presents a significant data analysis challenge. The manual vetting of stellar transit data—the primary method for detecting exoplanets—is a slow, labor-intensive bottleneck that limits the pace of discovery. This project's objective is to streamline this discovery pipeline by developing a reliable classification tool that can rapidly distinguish between genuine exoplanet candidates and false positives, enabling researchers to focus their efforts on the most promising signals.

Solution Architecture

The solution is a full-stack application architected with a clear separation of concerns:

- **Machine Learning Pipeline:** A classification model (e.g., Random Forest/XGBoost) is trained and validated on a cleaned and preprocessed version of the Kepler Objects of Interest (KOI) dataset. The final model, scaler, and feature set are serialized for production use.
- **Backend API:** A lightweight and scalable RESTful API is developed using **Flask**. It exposes a /predict endpoint that ingests new data points as JSON, preprocesses them using the saved scaler, and returns the model's classification with a corresponding confidence score.
- **Frontend Interface:** A responsive and intuitive single-page application built with **HTML, Bootstrap, and JavaScript**. The interface allows users to input key transit parameters, which are sent to the backend via an asynchronous fetch request. The returned prediction is then dynamically displayed on the page.

Impact and Significance

The primary impact of this system is the significant **acceleration of exoplanet candidate vetting**. By automating the initial classification, it acts as a force multiplier for researchers, allowing a greater volume of data to be analyzed with higher efficiency. The project's architecture is scalable to accommodate larger datasets from ongoing and future missions (e.g., TESS). As a portfolio piece, this project demonstrates a comprehensive skill set in full-stack development, data science, and foundational MLOps principles, including model serialization and API deployment.