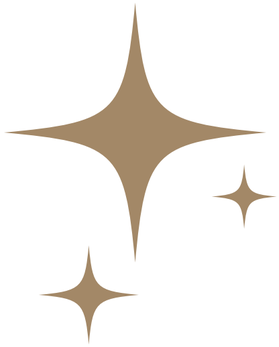
ECHO - Exoplanet Classification Hub

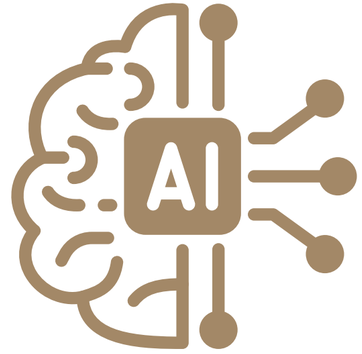
Participants: Ana Paula Serafim de Góis, Diego Anestor Coutinho and Lara Conte Gomes

Challenge: A World Away: Hunting for Exoplanets with AI



Link for the challenge→ https://www.spaceappschallenge.org/2025/challenges/a-world-away-hunting-for-exoplanets-with-ai/

The Exoplanet AI Challenge



Our challenge is to create an Artificial Intelligence/Machine Learning (AI/ML) model trained on one or more of NASA's open-source exoplanet datasets (from missions like Kepler, K2, and TESS).

The goal is not just to analyze this data to identify new exoplanets, but also to build an interactive web interface to facilitate user interaction.

It is crucial to deeply consider how the various data variables (such as orbital period, transit duration, and planetary radius) influence the model's decision to classify a data point as a confirmed exoplanet, a candidate, or a false positive. We recognize that strategic data processing, removal, and incorporation are fundamental to achieving higher-accuracy models.

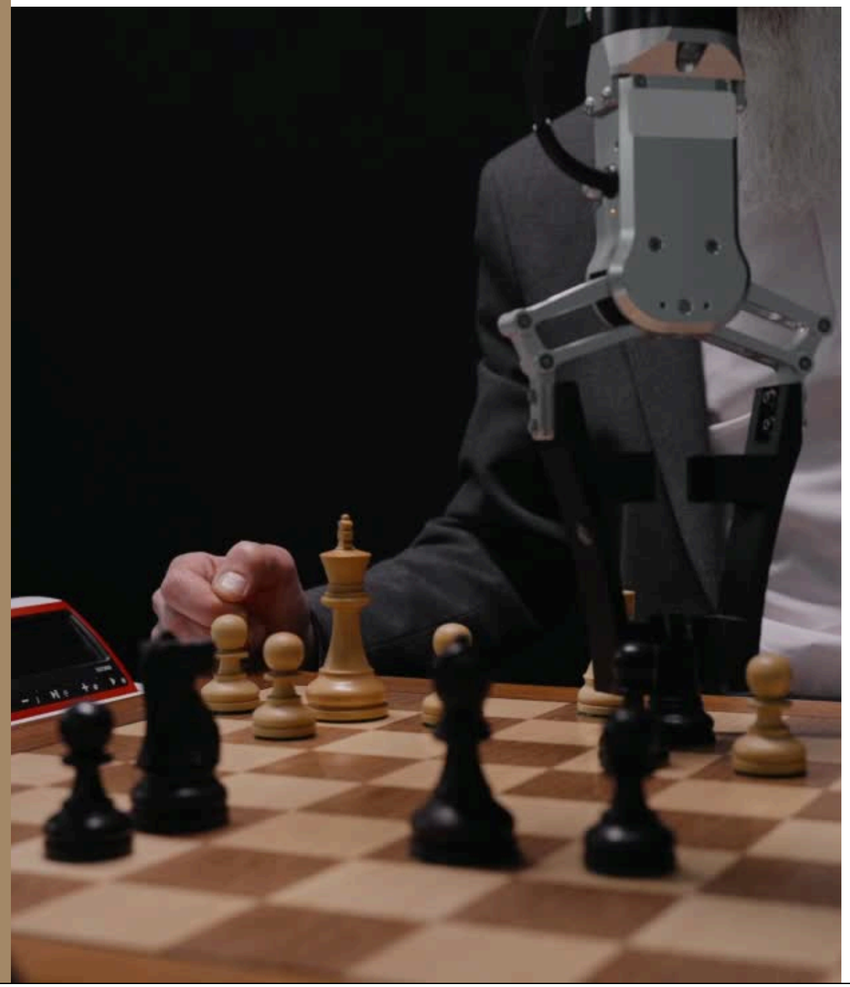
Finally, we must design the user interaction. We will decide whether to allow users (such as scientists and researchers) to upload or manually enter new data via the interface, and whether we will use that user-provided data to update and improve the model over time.

What data do we use?

- Sources merged: NASA Exoplanet Archive tables (TOI, Kepler/K2 cumulative, K2/Kepler candidates).
- IDs preserved for lookup: TIC, KEPID/KOI, hostname (kept exactly so we can fetch light curves from MAST).
- Ephemerides kept: orbital period, transit epoch (BJD), duration, and depth — the minimum needed to fold light curves.
- Catalog features normalized: stellar and system fields (e.g., st_teff, st_logg, st_rad, sy_dist, etc.).
- Unified schema: column names cleaned to lowercase, snake_case; duplicates removed; obviously empty columns dropped.

- Code: Add repo link here→https://github.com/DIEGOVZK/NSAC2025/tree/main/documentation
- Code: Add repo link here→https://github.com/apsgois/NSAC2025-ML

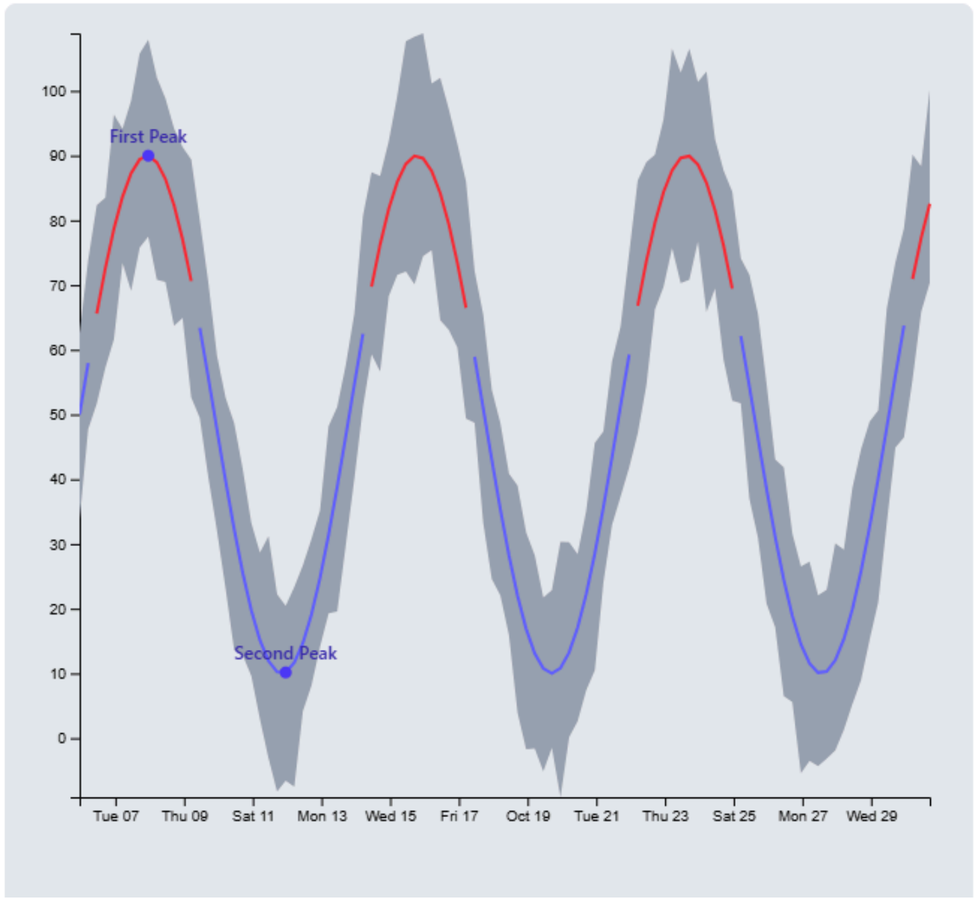
From Light Curves to Candidates: A Fast Pipeline

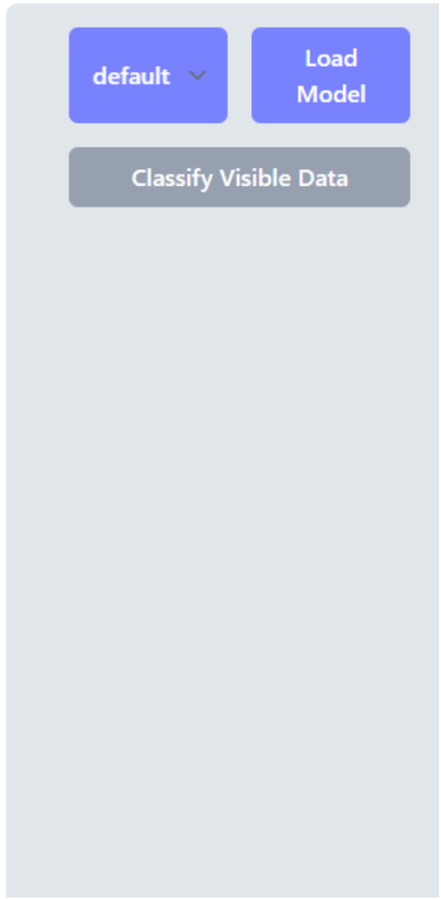


- Goal: Fast triage of TESS light curves with low compute, no training.
- Input: Text file of targets (TIC/TOI/name); optional sector filter.
- Pre-processing: NaN/outlier removal, flatten (detrend), median-to-1 normalization.
- Search: BLS (default) for periodic box dips; optional TLS with auto-fallback to BLS.
- Diagnostics: Period, t_0, duration, depth, SDE, SNR, #transits, plus odd–even and V-shape checks.
- Scoring: Simple rule-based planet_score (0–1) combining SDE/SNR and penalizing odd–even/V-shape.
- Outputs: scored.csv (all metrics + score) and Top-N plots (time + phase-folded with binned median).

- Code: Add repo link here→https://github.com/DIEGOVZK/NSAC2025/tree/main/documentation
- Code: Add repo link here→https://github.com/apsgois/NSAC2025-ML

Our website:





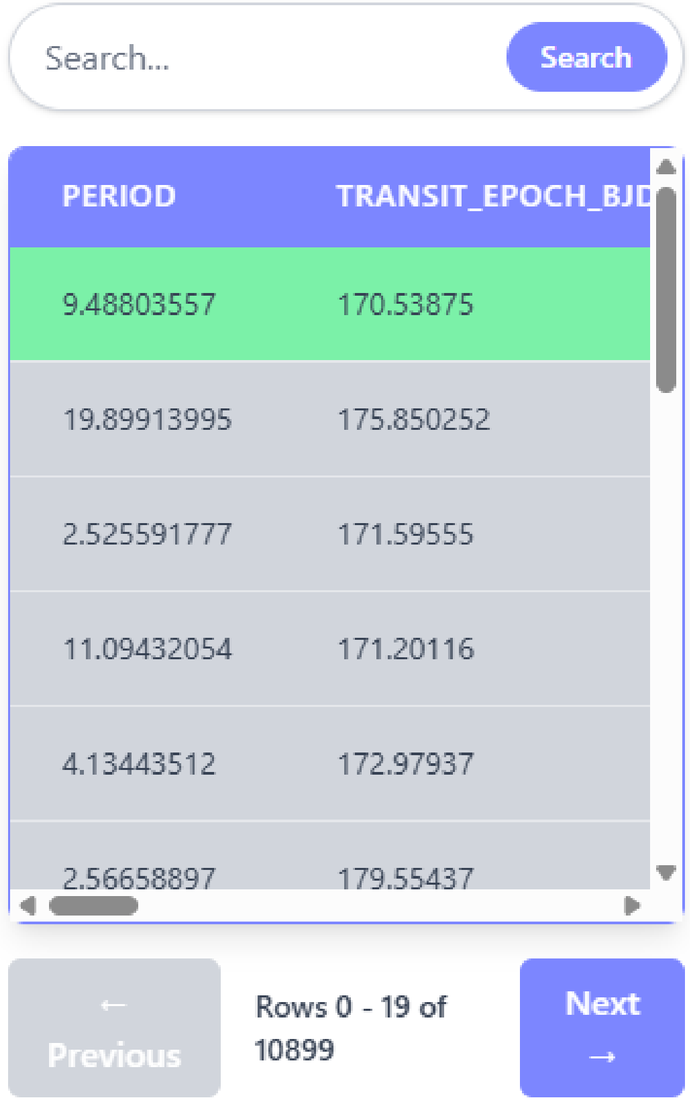
Website: https://exoplanetclassificationhub.study/dashboard

How it works

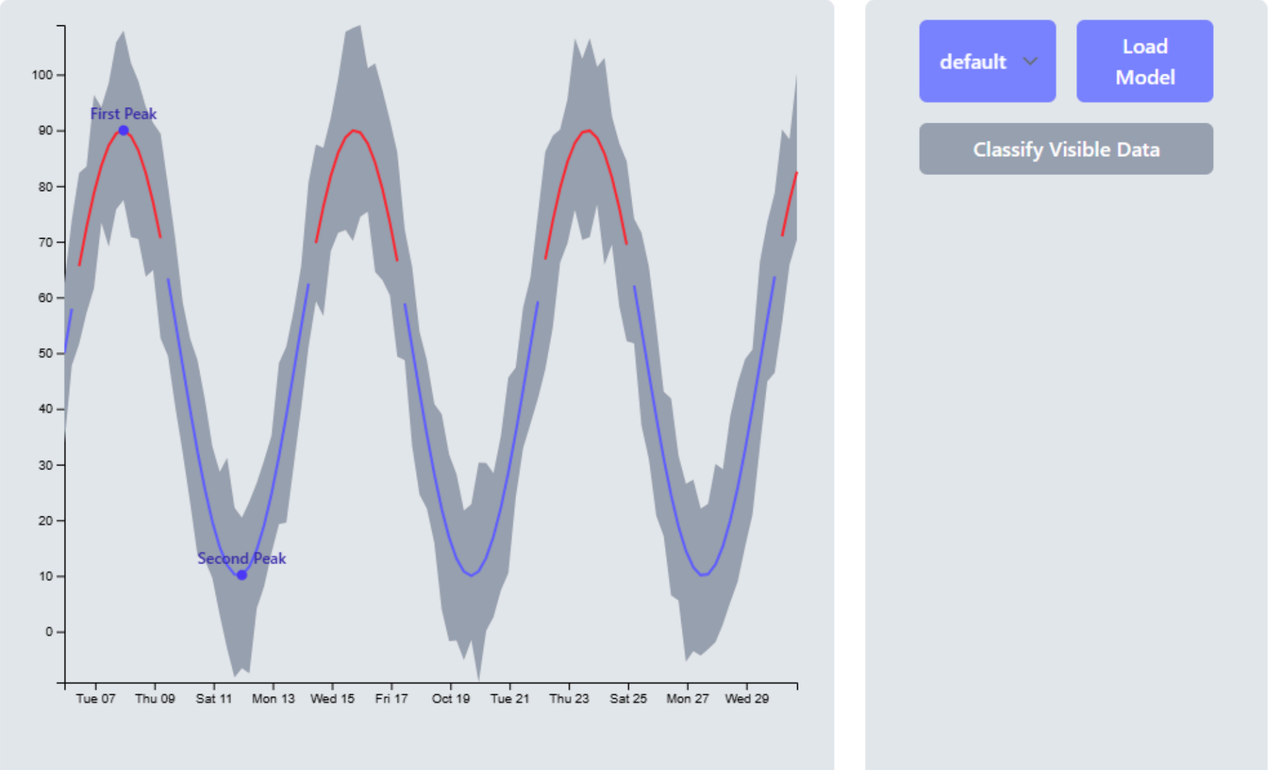
This section shows a table connected to the NASA Archive dataset.
From this table, we can select which data will be analyzed in the light curve.
After running the model, the table colors change to indicate:

- 🟢 True Positive
- 🔴 False Positive
- 🔵 True Negative
- 🟡 False Negative

When we select a row, the system automatically loads the raw light curve data related to that observation, allowing quick visual verification of the model's result.



Visualizing the Light Curve



Here, we will generate and display the simulated Light Curve Graph. This graphic is crucial as it visually represents the core evidence for an exoplanet: the periodic dip in starlight (the transit depth) caused by the planet passing in front of its star.

This element, located on the right side of the slide, represents the interactive visualization of our Machine Learning model and the exoplanet classification results.
It serves as the central hub of the web interface where users interact with the AI, feeding it data, seeing the evidence (the Light Curve Graph), and observing the model's final prediction (Confirmed Exoplanet, Candidate, or False Positive) in real-time.