Exoplanet Candidate Scoring and Target-Level Diagnostics

A physics-informed, calibrated ensemble for TOI, KOI, and K2

Ojas Gaikwad et al.

October 5, 2025

Challenge

Goal. Train an AI/ML model on one or more NASA open exoplanet catalogs to analyze new objects and identify true exoplanets.

Deliverables in our tool (`exo_analysis.py`):

- Calibrated probability $p \in [0, 1]$ that a candidate is an exoplanet.
- Target-level diagnostics: standardized features, local feature contributions, Monte–Carlo uncertainty, PCA neighborhood.
- Dataset-level validation: ROC/PR, reliability, threshold tables, feature importance.
- Works separately on TOI, KOI, K2; supports positive-only & PU regimes.

Data & Labels

Datasets. NASA Exoplanet Archive tables:

- **TOI** (many labeled: PC/CP vs FP) $\Rightarrow$ supervised.
- **KOI** (often positive-only) $\Rightarrow$ one-class/positive-only calibration.
- **K2** (candidates + meta) $\Rightarrow$ supervised/PU depending on availability.

Label map.

$$y = \begin{cases} 1 & \text{confirmed/candidate (PC/CP/CANDIDATE)} \\ 0 & \text{false positive (FP/false positive)} \\ \varnothing & \text{unlabeled} \end{cases}$$

Preprocessing. Comment/header stripping, robust parsing; median imputation + missingness flags; robust standardization via median/IQR.

Physics-Informed Features

We derive stable, dimensionless features from transit geometry and radiative scaling.

$$f_d = \frac{\text{depth (ppm)}}{10^6} \qquad \frac{R_p}{R_\star} \approx \sqrt{f_d}$$

$$\alpha_R = \frac{\text{Dur}}{P} \qquad \frac{a}{R_\star} \propto \frac{P^{2/3} M_\star^{1/3}}{R_\star}$$

$$S \propto \frac{R_\star^2 T_{\text{eff}}^4}{a^2} \qquad T_{\text{eq}} \approx \frac{T_{\text{eff}}}{\sqrt{2}} \sqrt{\frac{R_\star}{a}}$$

Plus: magnitudes (V/G/K), SNR proxies, impact parameter, eccentricity (when available), flags.
Standardization:

$$z_j = \frac{x_j - \text{med}(x_j)}{\text{IQR}(x_j) + \epsilon}.$$

Learning Regimes

Supervised (TOI)

Stratified CV with class weights. Unlabeled rows are held out for scoring.

Positive-Only (typical KOI)

Fit a *stack* vs. synthetic background; calibrate using internal CV (folds act as pseudo-negatives) $\Rightarrow$ reliable ranking and probability monotonicity.

PU Bagging (optional)

Sample pseudo-negatives from unlabeled; fit B base models; average posteriors:

$$\hat{p}_{\mathrm{PU}}(\boldsymbol{z}) = \frac{1}{B}\sum_{b=1}^{B} f_b(\boldsymbol{z}).$$

Stacked Ensemble & Calibration

Base learners: LR (elastic/weighted), Histogram GBoost, RF, SVM (RBF + Platt).

$$
\begin{aligned}
\{f^{(m)}(\boldsymbol{z})\}_{m=1}^{M} &\to \text{OOF meta-features} \\
s(\boldsymbol{z}) &= \sum_{m} w_m f^{(m)}(\boldsymbol{z}) + b \\
\hat{p} &= \begin{cases} \sigma(as + c) & \text{(Platt)} \\ \mathrm{Iso}(s) & \text{(Isotonic)} \end{cases}
\end{aligned}
$$

Reliability: Expected calibration error

$$
\mathrm{ECE} = \sum_{k} \frac{|B_k|}{N} \left|\mathrm{acc}(B_k) - \mathrm{conf}(B_k)\right|.
$$

We pick the calibration with lower Brier/ECE.

Target-Level Uncertainty & Explainability

Uncertainty

$$H = -p \ln p - (1-p)\ln(1-p), \qquad m = p - 0.5.$$

Monte–Carlo perturbations: $z_j \mapsto z_j + \eta_j$ with $\eta_j \sim \mathcal{N}(0, \sigma_j^2)$, $\sigma_j = \lambda\,\mathrm{IQR}(z_j)$; summarize $\{\hat{p}^{(r)}\}$ by mean, std, 5%–95%.

Local contributions Fit a ridge surrogate on k-NN in PCA space; rank $|\beta_j z_j|$ for the target.

Counterfactual proxy

$$\min_{\Delta \boldsymbol{z}} \ \|\Delta \boldsymbol{z}\|_2^2 \quad \text{s.t. } \sigma(\boldsymbol{\beta}^\top(\boldsymbol{z} + \Delta \boldsymbol{z}) + c) \geq \tau,$$

gives standardized "how-much-to-change" deltas.

Metrics & Thresholding

- ROC/PR AUC, Brier, ECE
- Threshold table: F_1, Youden J, balanced accuracy
- Class prevalence-aware summaries

$$\mathrm{AUC}_{\mathrm{ROC}} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\, d\mathrm{FPR}$$

$$\mathrm{AUC}_{\mathrm{PR}} = \int_0^1 \mathrm{Prec}(\mathrm{Rec})\, d\mathrm{Rec}$$

$$\mathrm{Brier} = \frac{1}{N}\sum_i (\hat{p}_i - y_i)^2$$

Outputs & Minimal Dashboard

Per dataset (TOI/KOI/K2):

- `dataset_predictions.csv`, `top_candidates.csv`
- `perf.csv`, `threshold_table.csv`, `feature_importance.csv`
- `fig/`: ROC, PR, reliability, PCA
- `targets/DESIG/`: PNG report + JSON reasoning
- `index.html`: minimal navigation to all artifacts

Target report panels:

1. Standardized feature bar chart
2. Local contribution bar chart
3. MC probability histogram
4. PCA neighborhood scatter (k-NN highlighted)

Computational Profile

- Robust z-scoring is $O(nd)$; imputation is linear.
- Histogram GBoost: $O(nd \log n)$; RF: $O(T\,nd)$; SVM: subsampled if n large.
- PU bagging $B = 25$ by default (tunable).
- Parallel CV folds; Monte–Carlo draws $R \in [128, 512]$ per target for fast uncertainty.

Scientific Sanity Checks

- **Transit consistency:** $(P, D, f_d) \rightarrow$ plausible $a/R_\star$, $R_p/R_\star$, T_{eq} for given T_{eff} and $R_\star$.
- **Calibration:** reliability curve near diagonal; small ECE.
- **Explainability:** top features align with astrophysical drivers (depth, duration, $R_\star$, T_{eff}).
- **Generalization:** dataset-wise training; KOI handled by positive-only calibration to avoid degenerate training.

Limitations & Extensions

Limitations

- No light-curve refitting (e.g., Mandel–Agol) in-loop; relies on catalog features.
- KOI positive-only cannot estimate absolute contamination without priors.

Extensions

- Incorporate light-curve fits for refined b, $R_p/R_\star$, SNR.
- Semi-supervised consistency regularization; domain adaptation across surveys.
- Bayesian calibration with astrophysical priors on size/period distributions.

How to Run

Full datasets

```
python exo_analysis.py ^
  --toi "C:\data\toi.csv" ^
  --koi "C:\data\koi.csv" ^
  --k2  "C:\data\k2.csv"  ^
  --out "C:\MATLAB\TOI_analysis\out"
```

First N targets per dataset First N targets per dataset

```
python exo_analysis.py --limit-targets 20 ... (same flags)
```

Outputs appear under:

$$\texttt{C:}\backslash\texttt{MATLAB}\backslash\texttt{TOI\_analysis}\backslash\texttt{out}\backslash\{\texttt{TOI,KOI,K2}\}$$

Takeaways

- Physics-informed features + diverse ensemble $\Rightarrow$ robust, calibrated probabilities.
- Target-level diagnostics (MC + local contributions) $\Rightarrow$ scientific interpretability.
- Works across TOI/KOI/K2 with supervision, positive-only, or PU bagging.
- Minimal dashboard + CSV/JSON artifacts for triage and paper-ready figures.