ExoVision

EXPLORE EXOPLANETS

Challenge

A World Away: Hunting for Exoplanets with AI

Member

GaHyun Jeong GyeongSu Kim Jonghun Shin
Sungphil Yang Hyundo Lee Chaerin Lee

NASA INTERNATIONAL® SPACE APPS CHALLENGE

“Seamless Interaction, Smarter Exoplanet Identification”

How can we explore and identify exoplanets more efficiently?
Our journey began with this simple yet profound question.

Today, most exoplanet detection still relies on manual analysis.
Despite the progress of AI and ML, data accessibility remains a significant challenge.

To enable automated discovery, we need more data — and to gather more data, we need broader participation.

So, what do we truly need?

1. **A platform where experts and non-experts can collaborate seamlessly**
2. **A new AI/ML model capable of analyzing exoplanet classification and probability**

Therefore, we set the following core missions:

1. **Build a precise environment for in-depth data analysis, enabling sharper insights into exoplanetary patterns.**
2. **Design an intuitive exploration experience, where anyone can navigate the cosmos with ease.**
3. **Harness the power of AI to push the boundaries of discovery and reveal worlds yet unseen.**

Unifying the Cosmos: From Data to Discovery

We combined **NASA's Kepler, K2, and TESS mission datasets** to create a unified foundation for exoplanet identification.
Our goal: **to transform fragmented observations into a single, learnable universe.**

Scientific Challenges

Inconsistent units, column names, and measurement methods

High proportion of missing values

Candidate labels represent uncertain states — potential planets yet unconfirmed

Why Kepler as the Core

1. Covers 45% of total data with 96% core features
2. 13 years of validated observations
3. Physically rich features suitable for machine learning
4. K2 and TESS data were refined and integrated into Kepler's schema

Reconstructing Missing Data

1. Physics-based reconstruction
   - Semi-major axis via Kepler's 3rd Law
   - Surface gravity via Newton's Law of Gravitation

   → **Restored 13,000+ missing values across 7 features**
2. Statistical estimation
   - Eccentricity inferred conditionally from orbital period

   → **Recovered 33,000+ missing values across 4 features**

Two-Stage Classification System

Model 1: CatBoost — 89.49% accuracy, CV ±0.55%

Model 2: Ensemble (CatBoost + XGBoost + LightGBM) — 74.39% accuracy

Integrated System: 70.86% (3-class), Overfitting: 1.04%p

Key Achievements

1. nified 3 NASA missions into one coherent dataset (21,271 samples, 25 features)
2. Reduced missing data by 77.8% using physics and inference
3. Built a two-model architecture with stable generalization
4. Achieved operational-level precision in binary classification

Scientific Insights

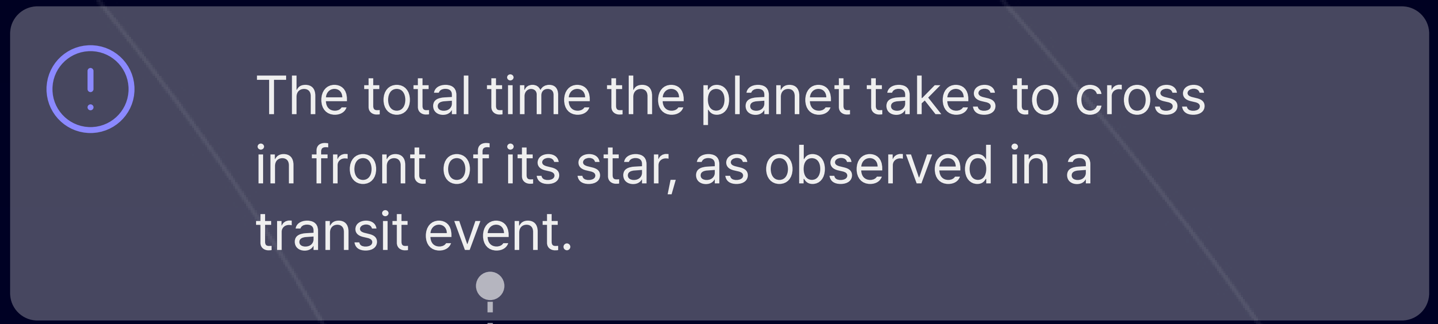
1. lanet radius is the strongest discriminator (13.68% importance)
2. Observation bias affects light-curve reliability
3. Metallicity correlates with gas giant probability
4. Candidate class inherently uncertain — requires further observation

Future Horizons

1. Advanced feature engineering and Bayesian tuning
   → **~92% accuracy**
2. Deep learning integration (CNNs, Transformers) on light-curve data
   → **95% potential**
3. Continuous TESS updates and ExoFOP validation
   → **stronger candidate classification**

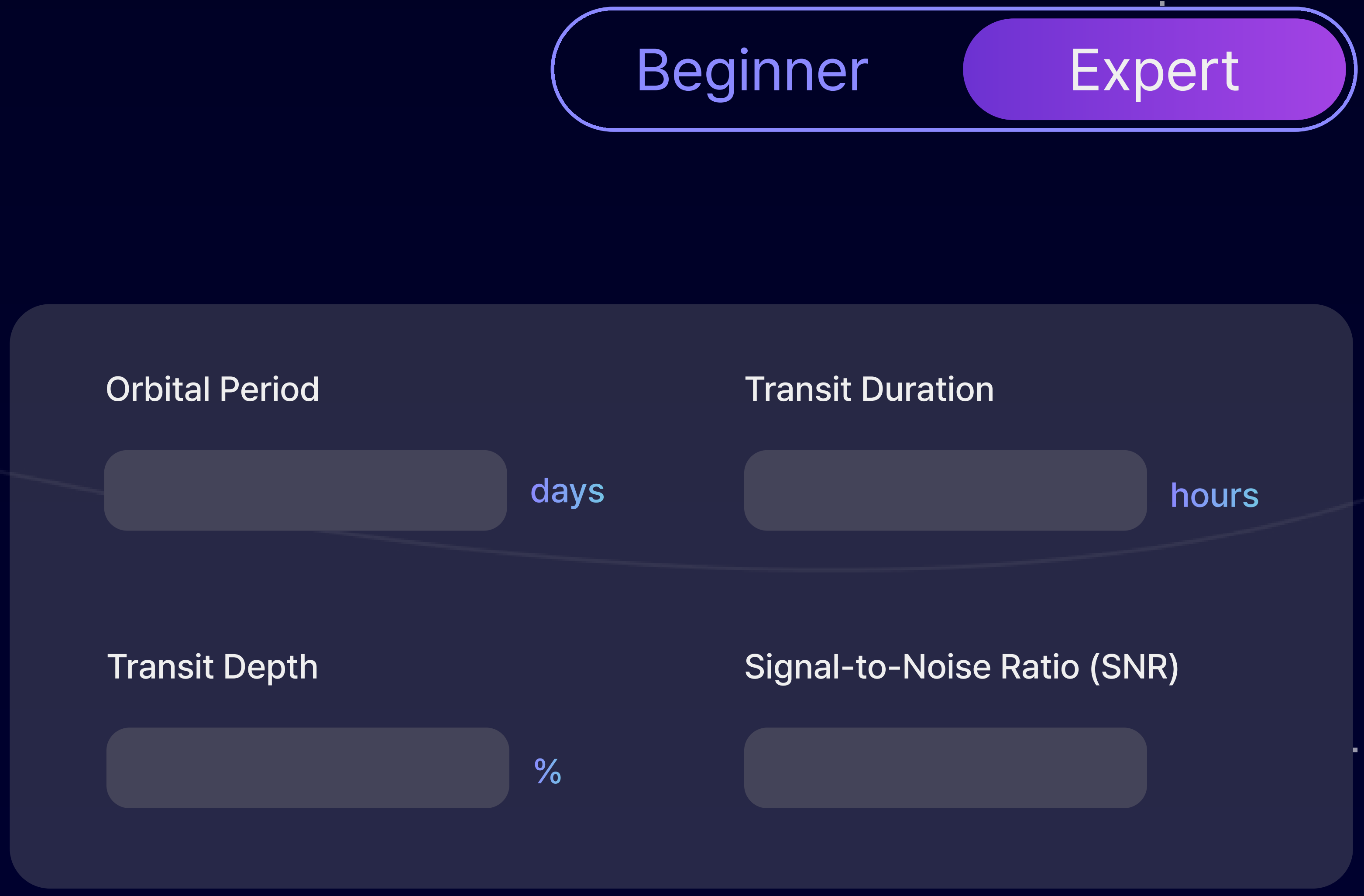
A simple toggle allows users to seamlessly switch between Beginner and Expert modes.

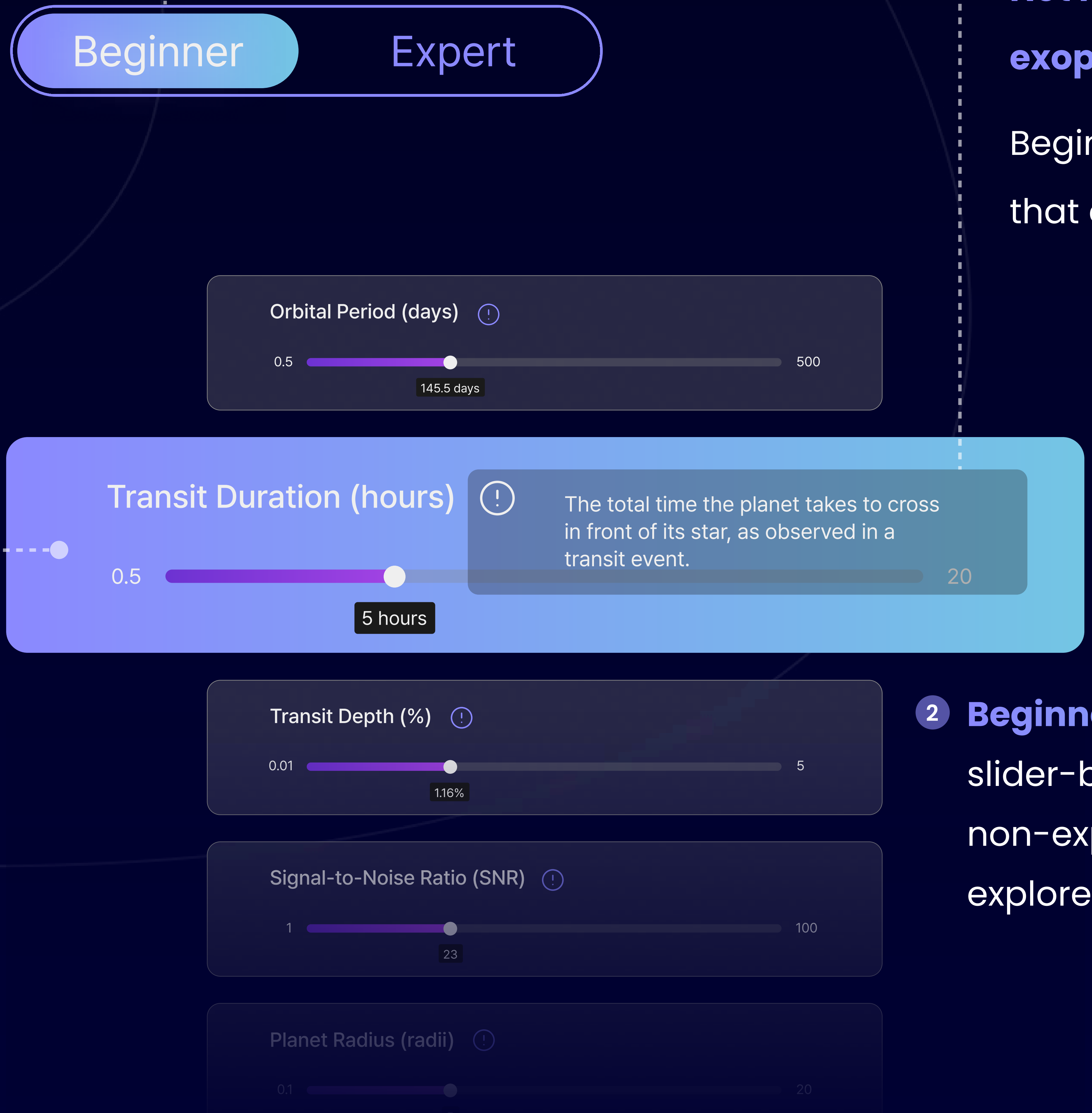
Exo Vision welcomes interaction from everyone — **from professional researchers to citizen scientists.**



Not familiar with the terminology used in exoplanet exploration?

Beginner Mode includes helpful **tooltips** that explain each data column in simple terms.





1 **Expert Mode** enables users to input exact numerical values for each data column, ensuring precision and control.

2 **Beginner Mode** features an intuitive, slider-based interface that allows non-experts to easily navigate and explore exoplanet characteristics.