

# Exoplanet Classification using NASA TESS Data

- ▶ Machine Learning for Space Discovery – NASA Space Apps Challenge
- ▶ Team **Lone Orbit** – COMSATS University Islamabad
- ▶ Project: **StellarSynth**
- ▶ Exploring planets beyond our solar system using NASA's data

The screenshot displays the 'Enter Observation Parameters' section of the StellarSynth application. It is divided into two main columns: 'PLANET DATA' and 'STAR DATA'. The 'PLANET DATA' column contains 'Planetary Properties' with input fields for Orbital Period (12.34 days), Transit Duration (3.50 hours), Transit Depth (0.001200), Planet Radius (1.10 Earth radii), and Insolation Flux (250.60). The 'STAR DATA' column contains 'Stellar Properties' with input fields for Star Magnitude (10.30), Star Temperature (5500.0 Kelvin), and Star Radius (0.90 Solar radii). Each input field has a numeric value and a range indicator (- +). The bottom of the interface shows the text 'CLASSIFY EXOPLANET' and a 'Manage app' link.

PLANET DATA	STAR DATA
<b>Enter Observation Parameters</b>	
<b>Planetary Properties</b>	<b>Stellar Properties</b>
Orbital Period (Days for one orbit)	Star Magnitude (Brightness value)
12.34	10.30
Transit Duration (Hours of transit)	Star Temperature (Kelvin)
3.50	5500.0
Transit Depth (Brightness decrease)	Star Radius (Solar radii units)
0.001200	0.90
Planet Radius (Earth radii units)	
1.10	
Insolation Flux (Energy received)	
250.60	
CLASSIFY EXOPLANET	
Manage app	

# Why Classify Exoplanets?

- Exoplanets orbit stars beyond our solar system.
- NASA's data includes Confirmed, Candidate, and False Positive planets.
- Manual classification is complex and time-consuming.
- Goal: Build a model to classify exoplanets automatically.

# NASA TESS Data Overview

- Datasets: TESS from NASA Exoplanet Archive.
- Contains ~7699 entries and 75 attributes (9 important ones).
- Checked duplicates, handled missing values, encoded categorical labels.
- Clean, structured data ready for machine learning.

# Our AI Pipeline



- Models Tried: Logistic Regression, Random Forest, XGBoost, SVM.
- Model Chosen: XGBoost
- Steps: Preprocessing, Train-Test Split (80/20), Label Encoding.
- Trained and evaluated models for accuracy and F1-score.
- Pipeline ensures consistency and reproducibility.

# Model Evaluation Results



- XGBoost achieved 68.1% accuracy (best performance).
- Random Forest achieved 67.5% accuracy.
- Other models performed moderately well.
- Metrics compared: Accuracy, F1-score.

# AI Empowering Space Discovery

- XGBoost identified subtle patterns in exoplanet data.
- Helps NASA automate and prioritize new candidates.
- Demonstrates AI potential in astronomical research.
- Turning Data into Discovery.

# Beyond This Mission

- Integrate JWST and future telescope datasets.
- Further enhance the interactive web-based dashboard.
- Extend to deep learning for higher precision.
- AI Helping Humanity Explore Beyond!